/c

MEASUREMENT
AND
STATISTICS
IN
EDUCATION

D. S. RAWAT

BUREAU OF EDNL. & PSYL. RESEARCH : LIBRARY

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

RAM PRASAD AND SONS : AGRA-3

(Binding)

S.C.E.R.T.. West Bengal
Date 25..4..66.
Acc. No. 1929

371.26
RAW

Thoroughly Revised & Enlarged
Third Edition 1965

© Author 1965

Price : Rupees Eleven Only
Agra Akhbar Press, Agra

PREFACE TO THE THIRD EDITION

The reception that this book received in its first and second edition has been sufficiently gratifying and favourable, and the comments that have come back have been sufficiently kind, that the author has tried to make this revision more elaborate and up-to-date.

During this span of eight years between the second and the third edition of the book, the author had a chance to visit U. K. and U. S. A. and get the latest knowledge about the developments in the field of evaluation and measurement. In this edition he has made a sincere attempt to make the book reflect the most desirable and significant developments in measurement. Not a single chapter has remained unchanged. Several chapters have been entirely or almost entirely rewritten. For example chapters on Personality assessment, Attitude measurement, Measurement of interests etc. have been added in this edition.

The major difference between this edition and the previous ones lies in increased emphases rather than in a change in the point of view. Through the inclusion of many illustrative exercises and examples the book has been made more practical. One full chapter has been devoted to the summary of the thesis submitted by the author in part fulfilment of the degree of M. A. in Education (Birmingham University). This chapter may be helpful to those research workers who are interested in comparing the achievement of two groups of pupils belonging to two different cultures. The achievement test in general science constructed by the author has also been given in this edition.

In the statistics part several practical examples from the educational data obtained in a real situations have been added. Efforts have been made to explain the statistical concepts in a simpler and easier way.

The revision of this book would not have been possible without getting the new ideas through the discussions with competent authors and scholars of the field. To this end my grateful thanks are due to Prof. E. A. Peel, Professor of Education and Head of the Department, Birmingham University, England ; Prof. W. J. Sparrow, Deputy for Post Graduate Training, Birmingham University and Dr. D. A. Pidgeon, Visiting Professor, Teacher's College, Columbia University, whose valuable discussions and suggestions were helpful to me in enriching the matter in the present edition of the book.

I shall fail in my duty if I forget to thank my publishers, for the pains they have taken in printing and publishing the book to their best capacity.

AUTHOR

PREFACE TO THE SECOND EDITION

In writing this prefatory note to the second edition of the present volume, I may say that in spirit and arrangement it is much similar to the first edition. As regards its contents, it may be added, I have endeavoured to my utmost to make necessary addition so that in its present shape it may claim to be called a thoroughly revised and enlarged edition. That the opportunity for a second edition has arrived so soon is an eloquent testimony to the utility that the book has served for the student community. I do not have any pretensions to superiority of my venture over authors of distinguished merit like Garrett, Vernon, Guilford, Lindquist, McNemar, Edwards, Yule, Kendall and others but I do feel proud of having written the book in a readable, simple and lucid style so that an abstruse subject frightening students sometimes like a bugbear may be turned into a joyful companion. Statistics, as far as my experience goes, is very often loathsome to students not because they cannot understand it, but because they cannot enjoy it. My attempt has been to make the reading of this book sufficiently interesting. How far I have succeeded in my endeavour will depend upon the response which comes from the readers.

It may also be added that the present volume has been divided into two parts : The First Part deals with a discussion of the general principles of statistics, and the attention of the readers is drawn towards new chapters are 'Regression and prediction ; 'percentile rank'; 'The graphic representation of correlation', which have been additions to this volume and which will enhance the influence of the book. Part two contains twelve chapters on educational measurement. The need for this additional portion arose out of the insistent demand from my students who urged that a simplified study of educational measurement would certainly help them to grasp statistics and measurement in education more intelligently. It is no gain saying the fact that statistics in education without educational measurement has no opportunity while educational measurement without statistics has no utility. I shall be happy if my readers give me a frank opinion regarding my attempt to present the two together. The twelve chapters of the second part of the book deal with 'An introduction to measurement'; 'The theory of test construction'; 'The construction of school subject tests'; 'New-Type tests'; Advantages and defects of New Type examinations'; Diagnostic and Instructional tests'; 'Intelligence tests'; The reliability and validity of test scores'; 'Analyzing and improving tests; 'Cumulative Record'; 'Reliability and Test of significance' 'The Chi-Square test'.

It is obvious, therefore, that my desire to make the book useful for students of L. T., B. T., B. Ed., & M. A, (Psychology) will certainly stand realised In the discussion of all these chapters I do not claim my pretensions to originality, but I feel indebted to standard authors like Lindquist, Hawkes, Mann, Ross, Micheels, Karnes, Anastasi, Remmers, Thurstone, Adkins, Gulliksen and others on whom I have freely drawn. What I do claim is the feeling of delight which has come to me as a result of my presentment of complex material in a logical, systematic apd simplified manner so as to make it useful and enjoyable to students. I verily believe that my colleagues and students in training colleges will give the credit if they find it due. My thanks go out to Prof. R. P. Sharma, M. A. (Eng.), M. A. (Hist.), B. T, for rendering me, from time to time, necessary help when I required from him. I am highly obliged to my publisher, who has shown sufficient patience with my slowness. I could not write the book at a stretch and the present book went to the press only in instalments. I heartily commend the man for his patience and zeal for bringing out this edition. I thank fellow teachers and friends for information as to errors and suggestions for improvement I have tried to take the fullest advantage of those suggestions. I am highly indebted to Prof. H. M. Dutta, M. A., M. Ed., A. I. E. (London), In-charge, B. R. College of Education, whose valuable discussions and suggestions are helpful to me in enriching the matter of the present volume to a large extent.

I may, in the end, mention to all my readers that I shall always be looking forward to welcome the opportunity if it comes from them by way of suggestions for further improvement.

AUTHOR

CONTENTS

CHAPTERS PAGES

PART I

MEASUREMENT

1. An Introduction to Psychological and Educational Measurement ... 3
2. Development of Measurement ... 14
3. Characteristics of a Good Test ... 26
4. Essay-Type Examiation *Vs.* New-Type Tests ... 43
5. A Programme for Constructing and Standardizing Achievement Test in any School Subject ... 62
6. Diagnostic Tests and Remedial Teaching ... 173
7. Intelligence and its Nature and Nurture ... 184
8. Psychology of Aptitude and Aptitude Tests ... 232
9. Attitude and Attitude Tests ... 241
10. Personality and Assessment of Personality ... 253
11. Analyzing and Improving Tests ... 276
12. Cumulative Record Cards ... 287

PART II

STATISTICS

13. Introduction ... 303
14. The Frequency Distribution ... 309
15. Averages and other Measures of Location ... 318
16. Measures of Variability ... 342
17. Correlation ... 368
18. Regression and Prediction ... 402
19. Graphic Methods ... 412
20. The Normal Probability Curve ... 434
21. Application of the Normal Probability Curve ... 452
22. Sampling ... 465
23. Reliability and Tests of Significance ... 470
24. The χ^2 (Chi-Square) Test of Significance ... 508

Table of Square and Square-Roots ... 525

Bibliography ... 535

CHAPTER 1

AN INTRODUCTION TO PSYCHOLOGICAL AND EDUCATIONAL MEASUREMENT

Measurement : a necessity of life

Measure is no rare freak. It gets up with us in the morning and goes to bed with us at night. Not only that but from birth to death almost every aspect of our daily lives is touched by measurement in its numerous form.

A baby is born and the nurse looks at the watch to record [illegible]e time. His weight is measured by a balance. He is feeded by a measured bottle. The child grows passing year by year. He goes to a school and observes his daily routine in a certain standard. He dresses himself in clothing of standard size. At the time of marriage he chooses a groom of a certain standard that may suit him. Perhaps now he becomes the father and begins to economise his expenditure, takes an insurance policy for a certain amount to meet the expenditure for the education of his sons. One day after another, year in and year out, he keeps this up until he may finally worry himself to death over a falling stock market, measured in rupees, and a rising blood pressure, expressed in points. He is covered by a cloth of his size and the time of his funeral is set according to the calendar and the clock.

So every time we are measuring something or the other, in one form or the other. But this all is the physical measurement. Psychological and educational measurement is more complex than physical measurement. It involves the mental processes of the individual which are not visible and which are interpreted in terms of the behaviour of the individual in certain situations.

Need for the psychological and educational measurement

Let us start by asking ourselves as to what is psychology and what is education. Psychology may be defined as the science of behaviour and experience. Behaviour is just what is observed. Much of modern psychology is concerned with these observations of behaviour and with the induction of laws from them. Education is the designed process aimed at developing an individual. "Its purpose is to promote the development of a well integrated person, capable of exercising such responsibility in society as his powers allow."[1] The word 'development' implies mental and emotional growth. Mental and emotional growth involves learning, thinking, perception and personal maturation. On the reverse side there is abnormal and retarded development. The end-product of this process of development is integrated person'. Educationists need to study psychology of personality. Individuals also differ in their endowments. It is, therefore, necessary to study the mental differences of persons both in intellect and temperament. We shall need also to know something of children's attainments and aptitudes. 'Responsibility in society' implies the attitudes of the individual to society, school family etc. As an educator we need to know something about the attitudes of the individual towards society, school, etc. In this book separate chapters have been devoted to the measurement attainment, intelligence, aptitude, personality and attitude.

Now if the purpose of education is to promote the development of a well integrated person, teachers must start their teaching with a certain objective which may bring a certain change in the behaviour of the students. The teacher is not able to promote the development of the individual. Only by defining the objectives. There are many individual differences. Teacher's function is to see which objectives are suitable for a particular individual or a group of individuals. Psychology may help the teacher for bringing about the expected change in the behaviour of the student. Teachers need to provide sufficient learning situations both in the class-rooms and outside so that desired outcomes may be achieved. Now the question arises

---

[4] E. A. Peel : *Psychological Basis of Education*, Oliver and Boyd, (1956), London.

whether we have achieved the goal for which the objectives have been defined. Whether we are able to promote the development of the individual. Here comes the question of measurement.

Educational and psychological measurement is not a new idea. Teachers have always endeavoured to measure the results of their teaching efforts as indicated by the progress of their pupils towards desired educational goals. Many have been equally concerned about the need to diagnose and remedy revealed defects in instruction. However, only recently has any large degree of accuracy been injected into their methods of measurement and diagnosis. Measurement of progress, evaluation of efficiency of instruction and the accompanying attempts at diagnosis were largely a matter of personal observation and judgment on the part of the individual teacher as recently as two decades ago. Actually, the recent development of modern educational instruments of measurement and evaluation may be regarded as an extention and improvement of an old practice. The modern educational measuring instrument presents a surprisingly accurate picture of the course objectives as well as analysis of the underlying skills, knowledge, concepts, understandings, and other outcomes upon which accomplishment in different subject matter field depends. It points out weaknesses in learning and instructional procedures. It permits the establishment of specific and objective goals of achievement that are based upon the actual attainments of children under typical school conditions. Educational tests and the information resulting from their use in the class-room have come to be almost universally identified with good teaching practice. Today the professionally equipped teacher is expected to be well versed in their construction, selection, and use in the class-room.

Difference between physical and mental measurement

Physical measurement is absolute whereas mental measurement is relative. If we say that the height of a pole is 10 feet, it will mean 10 feet to whichever part of the world we take it and measure it. If there is another pole 20 feet high we can say that it is twice as high as the first pole. This is an example of absolute measurement. On the other hand, if we say that

Manju secures 40 marks in English and Meena secures 80 marks in the same subject, it does not mean that the achievement of the latter is twice as much as that of the former. We can only say that the scores of the latter are 40 points more than that of the former. This is an example of relative measurement. Physical measurement is always absolute because we have always to start from an arbitrary zero point in the scale. Mental measurement is always relative because there is no arbitrary zero point from which we start our measurement.

Types of measuring rods in psychology and education

The various types of tests employed in education may be classified on different bases and from different points of view. For our purpose let us classify them in terms of their functions, their forms and their administration.

Tests can be classified as, 'educational', 'intelligence', 'aptitude' and 'personality', according to their functions. 'Educational tests' have, as their primary function, the measurement of the results or effects of instruction and learning. They are termed as general achievement test when they are designed to express in terms of a single score a pupil's relative achievement in a given field of achievement. They are also used as diagnostic tests to inquire into the nature of the difficulties which a weak pupil is having with his school subjects so that proper remedial instructions may be applied. In all cases educational tests measure the acquired abilities. On the other hand 'intelligence tests' or psychological tests have, as their purpose, the measurement of pupil's intelligence or mental ability in a large degree without reference to what the pupil has learned either in or out of school. Psychological tests are designed to measure the inborn abilities. Aptitude tests are used to measure the aptitudes of the persons. Personality tests attempt to measure such intangible aspects of behaviour as attitudes, interests, and emotional adjustment.

According to form the tests can be classified as follows :

(a) Power and speed tests,
(b) Verbal, non-verbal and performance tests, and
(c) Omnibus form and sub-test form.

(a) Power and speed tests. Tests of power and speed are used both in educational and intelligence testing, although personality inventories seem to involve neither the power nor the speed concept.

A power test is one in which the items are arranged from the easiest to the most difficult and in which the testee is given as much time as he requires to answer as many items as he can. A speed test is normally composed of items of similar degree of difficulty, and the testee is limited to a set time for the test. In actual practice most tests combine speed and power, that is, the items are arranged in order of difficulty and a time limit is set on the test.

(b) Verbal, non-verbal and performance tests. Another classification, cross-cutting, educational, intelligence, aptitude and personality tests is that dependent on the degree to which words are used in test items and in pupil responses.

'Verbal tests', by far the most common, are ordinarily of the pencil-and-paper variety although they may be oral or may even require identification of physical objects and material presented. In any event, words are used by the pupils in attaching meaning to, in responding to, or both in comprehending and in responding to the test items. A test unless qualified or further described is ordinarily a verbal test of the pencil-and-paper type.

'Non-verbal tests', again of the pencil-and-paper or even oral variety, are those in which pupils do not use the words in attaching meaning to or in responding to test items. Tests involving the use solely of numbers, of graphical representations, or of three-dimensional objects and materials are of this type.

'Performance tests' are also non-verbal but they may require use of pencil and paper by the pupils in responding, they may require solely the manipulation of physical objects and materials, or they may require paper-and-pencil responses to physical objects and materials presented in certain ways. Such tests are commonly used with persons having serious language handicaps and in situations where certain types of skills are of greater importance than in verbalization ability.

Distinction between verbal and language tests and non-verbal and non-language tests should perhaps be made here.

A verbal test is necessarily a language test. But a test may involve language, either oral or written, in the instructions given to the pupils and nevertheless be non-verbal if the test itself and the pupil responses do not involve language. Since performance test must be non-language as well as non-verbal, and hence involve the giving of instructions in pantomime.

(c) Omnibus form and sub-test form. The omnibus form is that in which the testee proceeds through the test to a time limit set for the whole test. Sub-test form is divided in several smaller tests, each of which is timed separately and the items of which are separately arranged in order of difficulty.

Individual and group tests

According to administration the tests can be classified either as individual tests or as group tests. Mostly this distinction is made in case of intelligence tests. Individual intelligence tests are administered to only one person at a time. Such tests require a full attention of a trained examiner. Although the techniques for administering these tests are highly standardized, the examiner modifies the procedure in various ways according to the age, ability, and even sex of the pupil being tested.

Group intelligence tests are usually paper-and-pencil tests that can be administered to a large group of persons at the same time. Group tests are objective in the making, but are not sensitive to the children's emotional differences.

Functions of testing

Almost any task or action some men can and will do more skilfully than others. Since these individual differences are a principal problem in directing the activities of others, every leader, manager, teacher, social worker and physician must be able to identify them. The purpose of every psychological test is to detect differences between individuals. This knowledge serves many functions such as prediction, comparison, diagnosis and research.

Prediction. An attempt to predict underlies every use of testing whenever a test is given to two persons, it tells about some difference between their performances at this moment.

But the fact would be of no significance, would not be worth knowing, if from it one could not predict that these two people would differ in some future activity. Consider a test as abstract as flash recognition of letters. Some people recognize four letters in a brief instant ; others react to five. This difference intrigues the curiosity, but it is an unimportant fact until it is related to some behaviour it predicts. The psychologist interested in applications may be concerned with it because this behaviour has much in common with airplane recognition and perception in reading. The theoretical psychologist is less, concerned with immediate application of laboratory findings, but he too will consider individual differences in a predictive sense by inducing from them general laws about perception and recognition applicable to the future behaviour of the individual.

Prediction is helpful in pupil gradation and placement. Administrators, supervisors and teachers find the problem of pupil placement one of the most difficult situations they have to face. The indefinite lines of division between the grades and the wide overlapping of ability between grades reveal that the typical techniques of pupil classification now in use are extremely crude. This could scarcely be otherwise in view of the methods commonly used. The proper grade placement of pupils implies that in so far as possible individuals who are normal for their group should be placed together. This means that pupils who are approximately alike in their chronological age, their educational achievement and their physiological, mental, social and moral development should, where possible, be placed together for instructional purposes. Not all of these qualities lend themselves readily to objective measurement, but a number of them do, and within these limits the results of objective measurements should be used in determining the pupils placement in his group. Through the development of reliable grade and age norms, based upon the achievement of groups of children on standardized tests, a valuable instrument for the establishment of grade lines and for within-class grouping is made available.

Prediction is the ultimate justification for the achievement test used for grading in school. It is used for prediction in two ways, although both teacher and pupil may overlook this fact in their concern with the grade itself. The grade is warranted primarily because it predicts the pupils' ability to carry on some future activity ; if it does not predict that, grading is likely not to be worth the troubles it causes. Furthermore, test results should give the teacher a prediction that if she teaches another class by similar methods, she can anticipate similar results of instruction. Without its predictive function, grading would be an act of crying over spilt milk.

Prediction is illustrated by clinical use of tests. A clinician gives a test of emotional adjustment to a client and finds that his score is far from the norm. Giving the test and using the results are based upon the assumption that a typical answers to the test predict that he will behave abnormally at some time in the future, unless corrective measures be taken. Prediction nearly always refers to some eventual overt behaviour. The clinician would not be set to detecting emotional maladjustment if that were only some internal state of the subject, never to crop out. The significance of the entire problem hinges on the fact that answers given on the test permit one to predict deviant behaviour which should be forestalled.

Comparison. Since the earliest beginnings of group instruction, class-room teachers have wished to know just how their pupils have compared in attainment with other similar pupils and classes. Until standard tests were developed, it was particularly impossible to secure this information. Now the giving of standardized test in arithmetic, science, Hindi, spelling, or other school subjects makes fairly easy a comparison of the results from a class with the norms established for the subject and grade.

Comparison with other classes within the system in which the teacher is working, within the same building, and even between different sections of classes in charge of the same or different teachers can be made on a basis of objective norms that have been derived for the various tests. Another comparison that is even more useful is that between the attainment of a class at the beginning and the end of a term or a year's work, or at shorter intervals in the course of a term. Each of these comparisons has its own peculiar value in assisting the teacher to determine the relative attainment and progress of his class at a given time.

Diagnosis. Educational diagnosis implies the use of more or less, technical procedures designed to locate specific learning and instructional difficulties, and if possible to determine their causes. The diagnosis of difficulties underlying educational accomplishments undoubtedly constitutes the high point in the supervisory and instructional uses of educational tests. Deficiencies of a general nature are revealed and brought to light by general survey tests. Specific weaknesses and to a certain extent, causes of weaknesses are identified by the use of properly selected diagnostic tests. Practically all of the more exact types of diagnostic procedures, such as the location of defects in speech, hearing and vision, are dependent upon educational test results for their initial steps. The diagnosis must be very specific. It is not enough to discover that a child is unable to read silently. The exact nature of his handicap must be revealed before it is possible to undertake a remedial programme. The more specific the diagnostic information revealed, the more exactly the remedial material can be made to fit the need. Suppose it is found by diagnosis that the child is unable to add, but unless the exact point at which his mastery of addition breaks down can be determined by the diagnosis, teaching or remedial efforts are largely wasted. One of the outstanding reasons why more effective teaching and remedial work has not been done in certain fields is that no adequate analysis of basic skills can be made or has been made. Remedial or corrective teaching is the result of deliberate constructive effort by the teacher after the particular points of weakness in the instruction of the pupils have been revealed by the tests. Diagnosis of the emotional handicaps of the pupils is also the function of diagnostic tests. Most pupils backward in school, in one or more subjects, labour under emotional handicaps which may be partly the cause of their scholastic difficulties. Such cases must be diagnosed and proper remedial instructions must be given.

Research. Attention may be drawn to research as function of testing, even though research involves prediction and occasionally diagnosis. In the search for knowledge about human behaviour, many questions arise which have only remote practical importance but which deal with fundamentals of biological or social science. Tests have frequently been designed to solve research problems, even though these tests have no every day utility. One example of such a test was developed by Lewis. He played phonograph recordings of words back-wards, in order to study the way people learn to recognize these strange sound patterns. Such tasks, just because they are novel, frequently makes very good experimental test situations. Another type of research makes use of common tests but seeks scientific knowledge rather than application ; an example is the study of differences between racial groups, in which psychological tests have been prominent.

Difference between measurement and evaluation

In an earlier paragraph of this chapter it has been said that the purpose of education is to promote the development of well integrated person, capable of exercising such responsibility in society as his powers allow. It was then pointed out that for the right development teachers must start their teaching with certain objectives. The development is indicated by the expected outcomes in the behaviour of the children. Now it is true that paper-and-pencil tests can measure only a limited portion of the outcomes of instruction and types of pupil behaviour about which the teacher and other school officers need information. By testing a child in mathematics we can measure his mathematical ability and nothing else. We cannot have any idea about the personality, temperament, attitude and aptitude of the child by applying this single test. So this is a case of measurement which is a narrow term. The present view is that objective tests constitute probably the major type of evaluative instruments but that such other means of measurement as the interview, the questionnaire, the cumulative record and the case study have a very significant place in evaluation of pupil behaviour and achievement. The evaluation concept has also doubtless been stimulated by the recent attention of educators and psychologists to whole child and his behaviour. This tendency to consider the child as a whole, rather than as an individual whose behaviour and abilities can be catalogued into a number of compartments, places a definite responsibility on the user of tests and other instruments of evaluation for considering the child in this broad sense. It is through the application of the evaluation concept rather than of the narrower concept of measuring that this result is most effectively obtained. By this time it is clear that measurement is a narrow term and evaluation is a broad term. Evaluation is a far more inclusive concept than measurement. Measurement merely provides data needed for evaluation.

The eight-year-study of the Progressive Education Association, which include both elementary and secondary education,[2] and the three year study of the Commission on Teacher Education on the college level[3] are illustrations of an enlarged conception of evaluation. The committee sought to devise suitable instruments of measurements for outcomes—such as interests, attitudes, creativeness, and various aspects of thinking—less tangible than those measured by ordinary tests and examinations. It also utilized other types of data, such as anecdotal records, family histories, records of the pupil's activities and the like.

Eugene R. Smith, Ralpha W. Tyler and Evaluation Staff : *Appraising and Recording Students Progress*, p. 550, (1942), Harper & Bros., New York.

[3] E. Troyer Maurice and C. Robert Pace : *Evaluation in Teacher Education*, P. 368, (1944), American Council on Education, Washington.

CHAPTER 2

DEVELOPMENT OF MEASUREMENT

Introduction

From the very early times measurement and tests have played a prominent role in the human history. Although the fact that individuals differ widely in their capacities and abilities has been recognized for several thousand years and educational measurement made formal entrance to the schools as early as mediaeval times, relatively little progress in educational test was made until the nineteenth century. In this chapter a brief history of the development of measuring techniques has been discussed.

Early oral examinations

The first evidences of the oral examination are found in ancient literature. The story is told in the Old Testament (Judges 12 : 5–7) of the test the Glieadites gave to the enemy Ephraimites who wished to cross Jordan. When asked to pronounce the word "Shibboleth" the Ephraimites could answer only with "Shibboleth", whereas people of the friendly tribes could respond with the correct pronunciation. Forty thousand Ephraimites were killed because they failed to pass this objective test.[1] Socrates, in a method he made famous, subjected his pupils to exhaustive and searching questioning. We hail him as one of the greatest teachers of all times, and we still apply in our schools, willingly or unwillingly, his method of teaching and testing by questioning and cross-questioning.

Early written examinations

Written tests are probably of more recent origin than oral quizzes but even they date back many centuries. As early as 2200 B. C., China had an elaborate national system of examinations and testing. In China public officials were required to pass a test in archery long after fire arms had become the established means of warfare. It means that for Chinese education was not a preparation for participation in adult activities but an end in itself. Public officials were related on the basis of written tests which were very lengthy and severe. Confined in isolated cells for hours at a time, candidates were compelled to write lengthy papers of treatise on assigned topics.[2] This shows that for Chinese education was not related to life. Brubacher in his famous book writes "China's pursuit of educational aims regardless of their congruity with the circumstances of life led in the long run to considering education not as a preparation for participation in adult activities, but as an end in itself. Indeed, for centuries the ideally-educated man in China was the scholar. To be sure, the scholar was also the civil servant, but nevertheless in later centuries much if not most of what he had to learn was quite unrelated to civil affairs."[3]

---

[1] Norma V. Scheidemann : *The Earliest Recorded Objective Test,* School and Society, 20 : 702 ; June, 1, 1929.

Recognition of Individual Differences

Individual differences among people have long been recognized. Plato, nearly four centuries B. C., divided his ideal society into three classes of workers, protectors and rulers. He believed that persons suited to each class should receive education for the fullest development of their personalities. Quintillian, shortly after the start of the Christian era, wrote that masters should observe differences in ability and inclinations of the persons they instruct, for the "forms of mind are not less varied than those of bodies".[4] In ancient India caste system was established on the basis of individual differences. It was believed that persons suited to each caste should receive education appropriate to their inclinations and task in the society. Brahmins were meant to teach, Rajputs to fight, trade was shared by Vaishyas and Sudras were to do the menial work.

---

[2] W. A. P. Martin : *The Chinese : Their Education, Philosophy and Letters*, pp. 45-47 (1881), Harper and Brothers, New York.

[3] J. S. Brubacher : *Modern Philosophies of Education*, McGraw Hill (1947).

[4] William Boyd : *The History of Western Education*, p. 76 (1921), A and C. Black Ltd., London.

The first tests used for measurement of the results or outcomes of education were probably not unlike certain of the performance tests of today, at least to the extent that they measured physical performance and that they were not paper-and-pencil tests. Among various primitive tribes, in which the young men were taught to hunt, fish and fight, the imitation ceremonies prerequisite to their admission to the ranks of adult males tested knowledge of tribal customs, endurance, bravery, other skills and abilities thought necessary for tribal protection. For Spartans, immediate aim of education was to make soldiers. Boys were given training which was calculated to make them subject to command, to endure, labour to fight and conquer. Newly born children were examined and weaklings were exposed to die on the mountain side. In ancient Athens, the stress upon athletic and aesthetic development led to evaluation by means of games and contests and of reading, writing and singing ability.

From this brief introductory note it can be easily concluded that measurement is not a recent development in the history of mankind.

The history of Intelligence tests

In earlier centuries people had a very vague notion about intelligence though the term 'intelligence' or 'intelligent person' was freely used. It had no psychological significance. They thought intelligence to be opposite of improperly thought out actions and an intelligent person was that whose actions were of intelligent nature.

The first systematic experimentation on individual differences in behaviour arose from the accidental discovery of differences in reaction time among astronomers. In 1796, an observatory assistant at Greenwich named Kinnebrook was engaged in recording, with great precision, the instant when certain stars crossed the field of the telescope. When Kinnebrook s results were found to be consistently eight-tenths of a second later than the observations of his superior, the Astronomer Royal, he was thought to have been incompetent in his work and was discharged. It was not until twenty years later that more careful study showed that the differences between observers were the result of the different speeds with which they could respond to stimuli. Only gradually did such differences come to be recognized as significant facts about human nature, rather than annoying errors contaminating scientific work.

Germany and experimental psychology

An important event in the history of psychology was the establishment of the first experimental laboratory in psychology by Wilhelm Wundt at Leipzig in 1879. He was, however, primarily interested in the analysis of consciousness into elements in a manner analogous to that employed in atomic chemistry. His sole interest in measurement appeared to be confined to reaction times, and he was distinctly unsympathetic to the problem of individual differences ; but he influenced considerably the course of psychology, especially the work of German psychologists, such as Kraepelin, Ebbinghaus, and Meumann.

England and foundation of statistical methods

Sir Francis Galton was the first man who recognized a clear cut individual differences. His interest in differences among individuals was stimulated by Darwin's newly published theory of differences between species. Up to the end of the eighteenth century and in the early part of the nineteenth century faculty-psychology was very common. According to faculty-psychology—the oldest structural theory—the mind is composed of a number of faculties or powers, each relatively independent of the others. Since the middle ages the major faculties of knowing, feeling and willing have played a big part in the theories of the nature of the soul. The faculties of judgment, memory and imagination were also recognized, and the number of the separate faculties grew larger as time went on. At the end of the eighteenth century the Scottish philosopher Reid gave a list of thirty powers of the mind, including the faculties of imitation, pity, duty, judgment etc. The number of these powers increased to thirty-seven in the early part of the nineteenth century.

Next came in prominence, the theory of mental structure. Whereas 'faculties' were defined largely as the result of introspection and speculation, 'mental factors' were studied by experimental methods. The empirical study of the structure of the mind can be traced back to Galton, when he introduced the idea of correlating two sets of results. Correlation is the very kernel of the method of analysing mental ability into its factors. For determining the mental ability of the different persons various mental tests are applied to them. The test scores are then subjected to mathematical and statistical analysis in order to arrive at the underlying factors of mental ability. The process of analysis is called factor analysis and has formed a considerable part of psychological and psychometric researches carried out in Britain and America since Spearman first discussed intelligence in 1904 and gave us mental factors in 1912. Spearman developed his well-known two-factor theory of intelligence on the basis of statistical analysis. According to this theory it is believed that a child's score on any test could be accounted for completely by a certain amount of general ability together with an amount of the specific ability for the particular test.

England and theories of mental structure

Karl Pearson, a student of Galton, formulated the method now most commonly used for calculating correlation coefficient. Soon after that many psychologists began to try this method in their analysis technique. Spearman's two-factor theory could not have seen the light of the day had Pearson not formulated the method for determining the coefficient of correlation. Soon after this came in the psychological field the sampling theory of Thomson. In this theory Thomson assumed the mind to be made up of many independent powers. Any particular test or school examination samples some of these powers or bonds.

Not after very long Kurt put forward a view that factors are descriptive categories and has advanced a theory of the structure of mind, commonly known as 'Burt's hierarchical structure', in which a place is found for all types of mental ability, general, group and specific.

Which theory is recognized in England at present? British psychologists in general, and Thomson and his co-workers in particular, have returned to a point of view not unlike that of Galton and Pearson in assuming a general factor of intelligence combined with more specialized group factors. However, as a rough estimate, Spearman's two-factor theory is still useful.

France and origin of intelligence test

The French were the first persons who approached the problem of measuring intelligence from the stand point of the classification and treatment of mentally defective. Alfred Binet discovered, in 1905, that by testing a large number of children (selecting them at random) and taking the average performance of the whole group, it is possible to find out the average performance of a normal child. Those children whose scores tally with the average score of the whole group of the same age would be the normal or average children. Binet first studied law, then medicine, and afterwards worked in a biological laboratory. Later he turned into a psychologist. Binet earned the name of Father of Intelligence. The first intelligence scale was produced by Binet in collaboration with Simon in Paris. In 1908 the intelligence scale was issued in a revised form, and finally in 1911, the scale as we have it at present was published. Some tests which had been found unsuitable in the previous scale were dropped. Binet gave a significant idea of 'Mental Age'. For example it became clear that a ten years old child doing the tests of an eight years old child was retarded two years out of ten. A ten years old child able to do the tests expected of the normal twelve years old child was obviously of the same mental ability or 'mental age' as twelve years old boy.

Psychologists of other countries took much advantage of this new concept of mental age. In 1912 Stern represented the intelligence as a ratio of mental age to chronological age and thus introduced a new concept of Mental Quotient. Symbolically Mental Quotient $= \frac{\text{M. A.}}{\text{C. A.}} \times 100$. The ratio is multiplied by 100 to avoid decimal fractions. This concept for which Stern suggested the term 'Mental Quotient' was later adopted by Terman as the familiar I. Q.

Attempts to measure intelligence in America

Dr. E. S. Chaille, an American physician, is credited as early as 1887 with the development of standards and simple tests judging the mental levels of children to the age of three and with having implied, although not definitely used, the concept of mental age as index of mental maturity.

Cattell apparently first used the term 'mental test' in 1890, almost at the beginning of the period during which scientific method was first applied to the measurement of mental ability. Cattell worked on the lines of Galton. He attempted to measure intelligence, but with the wrong tools. Very little attention was given to reliability and validity. In 1901, when Wissler published his analysis of Cattell's tests used with college students, in which he applied for the first time the Pearson's correlation technique to test scores, and in which he found little more than chance relationship either among the tests or between the tests and college work, a considerable damper was thrown over the enthusiasm of American testers which was not lifted till after Binet had published his 1905 scale.

Goddard, Kuhlmann and Terman all attempted the Binet-Simon test use with American children during the period from 1911 to 1916. Terman and his collaborators made the 'Stanford Revision of the Binet Scale' available in 1916, and in 1937 followed it with a second and more complete revision. These revisions make use of the Intelligence Quotient (I. Q.), based on the relationship between a child's mental age and his chronological age.

In 1917 a pressing practical need for group intelligence tests was felt by the Americans. When the United States entered the First World War in 1917 a committee was appointed by the American Psychological Association to consider ways in which psychology might aid in the conduct of the war. This committee under the direction of Robert M. Yerkes, recognized the need for the rapid classification of million and a half recruits with respect to general intellectual level. Such information would be of assistance in many administrative decisions including rejection or discharge from military service, assignment to different types of services, or admission to officer training camps. It was in this setting that the first group intelligence test was developed. In this task, the army psychologists drew upon all available test materials and especially upon an unpublished group intelligence test prepared by Arthur S. Otis, which he turned over to the army.

The test finally developed by the army psychologists have come to be known as the 'Army Alpha' and the 'Army Beta.' The former was designed for general routine testing, the latter was a non-language scale employed with illiterates and with foreign-born recruits who were unable to take a test in English, The pattern established by these tests was closely followed in subsequent development of a large number of group tests for civilian applications.

At this stage it will not be out of place to mention the theory of mental structure that is common in America. In America a 'theory of primary mental traits' has been developed by Thurstone in 1935. Thurstone's analysis shows that the best way of interpreting the correlation data is not merely to suppose a 'g' ability and different 's' abilities, but to suppose a finite number of factors or abilities of which 'g' may be the first and most important. Thurstone finds that the number of mental factors or abilities that are of major importance in our daily lives are finite and may be briefly enumerated 'V'—Verbal factor ; 'N'—Number factor ; 'W'—Word fluency ; 'S'—Space ; 'M'—Associative memory ; 'P'—Perceptual speed ; 'R'—Reasoning.

Aptitude or specific intelligence tests

When a psychologist speaks of an aptitude he has in mind the potentiality a person has to succeed in an occupation or school attainment. So far we were talking about the measurement of general intelligence but persons possess some specific intelligence, or a capacity in restricted area, such as musical talent, clerical efficiency etc. To measure this specific intelligence we use the aptitude tests or prognosis tests.

In 1913 Münsterberg constructed tests for telephone girls and motormen. In 1915 Seashore Test of Musical Talent was constructed in America. In 1918 there appeared the Stenquist Test of General Mechanical Ability. In 1918, also, Rogers published a test of mathematical ability, which, although hardly an aptitude test in the modern sense, introduced the idea which other authors have followed up by aptitude tests in the special branches of mathematics, such as algebra and geometry. Spearman's splitting of total mental ability into a general factor and many specific factors had its influence on the development of aptitude tests for various abilities. Thurstone's attempt to devise tests of what he terms 'primary mental abilities' was also an incentive to aptitude test constructors.

A brief history of personality tests

From the very early times people began to rate persons for

Date 25.4.66
Acc. No. 1929

371-26
RAW
1929 .66
R.T., West B
Library

various emotional, personal and temperamental qualities, *e g.,* for narrow interest, joy, curiosity, emotional stability etc. No doubt these early attempts were unsystematic and unreliable but the modern techniques of measuring personality derived much from them.

Free association tests and projective techniques

Personality testing had its antecedents in the work of Kraepelin and Sommer on free association tests during the last decade of the nineteenth century. The modern projective techniques have been originated from these free association tests. In 1879 Galton developed the 'free word association' test. This is one of the earliest methods of assessing personality. Jung in 1905 published a free association test designed to reveal emotional complex. An entirely different approach to the quantification of word associations was put forward by Kent and Rosanoff in 1910-11. Pressey published his X—O or cross out test in 1921. Although the 'Roschach', the first modern projective test, was introduced in 1921, it was not until some thirty years ago that projective techniques employing such unstructured situations as inkblots and pictures came into wide use in the study of personality. Another projective technique which is widely used is the Thematic Apperception Test (T. A. T.) which was developed by Morgan and Murray in 1935.

Rating scales. The first rating scale in a modern sense was probably that of Galton for mental imagery, which was published in 1883. One of the most famous scale is Man-to Man Scale which was introduced in America during the first world war to rate the Army Officers for leadership. Freyd's 'Graphic Rating Scale' was published in 1923. Very soon after this many analytic scales were published by various psychologists. Among the famous scales are : Marton's scale (1925), Bridge's scale (1931), Laird's scale (1925). In 1934 Moreno's sociometric technique was published. A recent development in rating scales is the 'Rating within Persons'. This technique has been well explained by Burt and Stephenson in British Journal of Psychology (1937, 28, 59-96 and 1936, 26, 344-61).

The questionnaire. The questionnaire method used by Galton was the dominant early method of personality measurement in England. Galton, very soon, improved and used extensively the questionnaire which was imported to America about 1880 by G. Stanley Hall. The instrument exists in many forms today, but its principal use in education is for measuring adjustment, attitude and interest.

Most of the hundred or more tests that have been published are modifications or extensions of three prototypes: Woodworth's 'Personal Data Sheet' (1920), Frey-Heidbreder's 'Introversion-Extroversion Test' (1924-26), and Allport's 'A—S' 'Ascendance-Submission Reaction Study' (1928). It is recognized that attitude plays an important role in education. Since about 1920 much attention has been devoted to the measurement of attitudes of various kinds. Hart's test of social attitudes and interests which appeared in 1923 and Watson's measurement of fairmindedness which appeared in 1925 are good examples. Beginning in 1928 Thurstone has been responsible for important improvements in the units of measurement employed in attitude questionnaires on many subjects. In 1932 Likert published tests of Internationalism, Imperialism and Attitude to Negroes. He also published a new technique for the measurement of attitudes. Among other well known attitudes scales are: Remmer's 'Generalized Attitude Scale' (1934) and Bogardus's 'A Social Distance Scale' (1933).

Regarding the questionnaire for measuring personal adjustment mention may be made of 'Minnesota Multiphasic Personality Inventory' developed by Hathaway and Mckinley in 1940. Bernreuter's 'Bernreuter Personality Inventory' was developed in 1933. It claims to measure four personality traits—Neurotic tendency, Introversion-Extroversion, Dominance-Submissiveness, and Self-Sufficiency. People also tried to construct personality tests for young children. 'Maller's Character Sketches' (1932) is one of the best questionnaires for children.

Interview. The interview is the method most frequently used for assessing personality for educational and vocational purposes, and for the diagnosis of maladjusted children and mental patients. It is also one of the oldest method for obtaining the knowledge regarding the individual's personal adjustment. Interview method is not much reliable and lacks in validity. It can be used with advantage for supplementing other objective methods of assessing personality.

Achievement tests. It has already been pointed out that the early Chinese (about 200 B.C.) are credited with having used tests in the selection of their civil servants. Of course, their tests were neither objective nor standardized.

Written examinations in the schools are of more recent origin. Horace Mann suggested the use of written examinations in 1845. It is interesting to note that he advocated the use of a large number of questions and commented on the desirability of standardization. At about this time Boston schools began to administer written examinations. A few years later (1865 to 1878) the Regent's Examinations were inaugurated in the State of New York. In 1900 the College Entrance Examination Board was organized and the next year provided a few broad essay questions that were taken by approximately 1000 private school graduates who wanted to enter colleges cooperating with the Board. In 1864 an English Schoolmaster, the Reverend George Fisher, evidently realizing the subjectivity of ordinary examinations, proposed a 'Scale-Book', made up of 'various standard opinions......... arranged in order of merit'. But these efforts could not produce lasting effect. However, as soon as the use of written examinations become widespread, they began to be criticized. School superintendents as well as college administrators gave impetus to the growing discontent. Almost parallel to these subjective voices being raised was the emergence of the scientific movement in education. In 1890, F. Y. Edgeworth was writing in the 'Journal of the Royal Statistical Society' about "The Element of Chance in Competetive Examinations.". This article could show very competently and objectively that the existing examinations were unreliable.

The need for reform in college marking was forcibly brought to public attention by Max Meyer, who reported on the marks collected from forty instructors for a period of five years at the university of Missouri. He found such astonishing variations as 56 per cent of A's in Philosophy and only 1 per cent in Chemistry III, while there were 28 per cent of failures in English II and none in Latin I.

Thorndike or his students are credited for the first standardized tests. The first test was the Stone Arithmetic Test published in 1908 and the first scale was Thorndike Handwriting Scale announced in 1909 and published the following year. Soon after these other standardized tests made their appearance, and by 1928 the number had grown to almost 1300, as catalogued by Monroe and his associates. Several years ago Lee was writing, justly proud, about the 6 million standardized tests used in the schools in 1931. More recent figures show that during 1944 approximately 60 million standardized tests were administered to approximately 20 million people in America. Almost the same figures are for England.

Monroe has given a brief summary of the measurement movement for the quarter of a century beginning in 1920. It was during these eventful years that educational measurement passed from early adolescence to early adulthood.

CHAPTER 3

CHARACTERISTICS OF A GOOD TEST

What makes a good test ?

BEFORE DISCUSSING THE NATURE OF THE VARIOUS TESTS AND THEIR construction we should give some thoughts to the characteristics of a good test.

If you were told to construct a fine piece of furniture according to the highest standards, you would naturally be expected to know and understand those standards. If you were not sure of some of the requirements, you would want to find out about them right away. This is also a basic step in learning about various types of tests, how they are built and how they can be used. What makes a good test ? What are the characteristics and requirements that must be kept in mind ? These are the questions to be answered first.

If a general statement would suffice, these questions could be answered in one sentence—a good test is one that does what it is supposed to do. This can be further explained in the following manner :

1. A good test must measure accurately and consistently (reliability).
2 It must actually measure what it is supposed to measure (validity).
3. It must be fair to the students (objectivity).
4. It must pick out the good students from the poor (discrimination).
5. It must be long enough to do the job (comprehensive).
6. It must be easy to use (ease of administration and scoring).

All these factors should be present in a good test. They are interdependent. They affect each other. For purposes of discussion each of the factors will be considered separately.

but we should keep in mind that they are mutually causal and have direct bearing on each other.

The test must be reliable

A reliable test is a 'trustworthy' test. It is accurate. It is consistent. If the test measures exactly in the same manner each time it is administered, if the factors that affect the test scores affect them to the same extent every time the test is given, the test is said to be high in reliability. In other words, the reliability of a test, as of any measuring instrument, depends upon the consistency which it gauges the abilities of those to whom it has been applied. A highly reliable test should yield essentially the same score when administered twice to the same student, provided, of course, that no learning occurs while student, provided, of course, that no learning occurs while the test is being taken the first time or no learning or forgetting takes place between testings.

The following are the methods of determining test reliability :

1. The test-retest (repetition) method.
2. The alternate or parallel form method.
3. The split-half method.
4. The method of 'rational equivalence'.

1. The test retest method. This is the simplest method of determining the test reliability but it estimates the reliability in the test less accurately and hence it is used less generally. The procedure is quite simple. The test is administered to a certain group of testees and then repeated on the same group after a certain interval of time but under the same conditions. If the testee's scores differ widely from those previously obtained, the test is obviously a poor one. On the other hand, if there is a high correlation between the two sets of the time scores, the test is said to be reliable. This method is defective because the testees may remember their previous response and at the time of repetition of the test they may recall them and spend their time on new materials, thus increasing their scores. The repetition of a test may also be affected by the memory effect, practice and confidence induced by familiarity with test materials. If, however, the interval of time between the first and the second test be long (say six months or so) these effects can be lessened but then the growth and maturity of the subject (if they are children) will affect the test.

2. Alternate or parallel forms method. The alternate or parallel forms of test can be constructed and administered to the same group under similar conditions. These tests are called parallel in the sense that they are similar in their form, content and difficulty. The correlation between these two forms is taken as a measure of self correlation. This method is also not free from memory and practice effect. When the second test is administered after the first one, the subjects scores increase on account of these factors. In case these increases are constant (say four points for each score), the raliability coefficient of the test will not be affected. Both the tests may be called reliable in such cases but to compare the scores of the second test with the scores of the first one, we will have to determine the constant increase in each score due to practice and memory effect and then to subtract the constant amount from the scores of the second test. The parallel forms of the test should not be a pure duplicate, the more they are identical the higher will be the coefficient of reliablity and the less they are similar the lower will be the coefficient of reliability. This method is generally used in the psychological and educational tests where alternate forms are available.

3. The split-half method. This method is used when no alternate forms are available or when it is not feasible to repeat the same test. In this case we split the test in two parts. The correlation between the two halves is determined. From the half-test reliability, the self correlation of the whole test is detemined by the Spearman-Brown formula described in the later section of this chapter.

For splitting the test we can adopt different procedures. One may be the scores on odd—and on even—numbered question or items may be totalled separately, and the inter-correlated. The split-half method has also a few disadvantages. The test reliability depends upon the number of items, *i.e.* the length of the test, hence the correlation between one half and the other half is unduly low. It is corrected by a formula discussed in the later section of this chapter.

The chance error may affect the scores on both the halves of the test equally, thus our test may appear to be highly reliable. The last objection about the split-half method is that the test can be splitted into two halves by various methods and thus the reliability coefficient is not a unique value.

The split-half method is generally used because in this case we get the date for determining the reliability upon one occasion.

4. The method of 'rational equivalence'[1]. The defects of all the above methods have been eliminated by the use of the method of 'Rational Equivalence.' The two tests are called 'equivalent' when their corresponding items are inter-changeable. This method enables us to compute the inter correlations of the items of the tests and the correlations of the items with the test as a whole.

The most useful formula for this is

$$r_{11} = \frac{n}{(n-1)} \times \frac{\sigma_t^2 - \Sigma(pq)}{\sigma_t^2} \qquad (A)$$

where

r_{11} = reliability coefficient of the whole test.

n = Number of items in the test.

p = The proportion of the group giving the correct response of the test item.

q = The proportion of the group giving the wrong responses to the test item.

Since $p+q=1$

$\therefore \quad q=(1-p)$

σ_t = The standard deviation of the test scores.

Firstly we calculate the σ_t, *i.e.*, the standard deviation of the test scores for the whole group. Then the proportion of the group passing each item and failing each item is obtained. These we call as p and q respectively. The p and q are multiplied in each case and the sum of all the (pq)s is obtained which is known as $\Sigma(pq)$. The final step is to substitute these values in the above formula to obtain the reliability coefficient of the whole test. We may illustrate this with the help of an example. Say a test of 50 items has been administered to 100 (eleven years) boys. If $\sigma_t = 7{\cdot}5$ and $\Sigma(pq) = 19{\cdot}5$ we can use the above formula to calculate the reliability coefficient of the whole test.

---

[1] G. F. Kuder and Richardson M. W. : *The Theory of Test Reliability*, Psychometrika, (1937), 2, pp. 151-60.

$$r_{11}=\frac{n}{(n-1)}\times\frac{\sigma_t^2-\Sigma(pq)}{\sigma_1^2}$$

$$=\frac{50}{(50\text{-}1)}\times\frac{(7\cdot5)^2-19\cdot5}{(7\cdot5)^2}$$

$$=\frac{50}{49}\times\frac{36\cdot75}{56\cdot25}$$

$$=1\cdot02\times\cdot65$$

$$=\cdot66$$

This ·66 is the reliability coefficient of the whole test.

Sometimes for quick computation work use the following formula to determine the coefficient of reliability of the whole test.

$$r_{11}=n\,\frac{\sigma t-m(n-m)}{\sigma_t^2(n-1)} \qquad \text{(B)}$$

where

r_{11}=Coefficient of reliability of the whole test.
σ_t=Standard deviation of the test scores.

n=Number of items in the test.
m=The mean of the test scores.

This formula is used in the case of such tests in which the test items have the same degree of difficulty. The same proportion of the subjects may be able to pass each item. It is not necessary that the same persons may pass each item. The greater is the variability in the difficulty of the test items the larger the under estimation of reliability is got by this formula.

To illustrate, suppose a matching test (30 items) has been administered to a group of fourteen year boys. An item answered correctly is scored one, an item answered wrongly is scored zero. The mean test score $(m)=20$ and $\sigma_t=5$. What is the reliability coefficient of the test ?

$$r_{11} = \frac{n\sigma_t^2 - m(n-m)}{\sigma_t^2(n-1)}$$

Substituting the given values in this formula :

$$r_{11} = \frac{30 \times 25 - 20\ (30-20)}{25 \times (29)}$$

$$= \frac{30 \times 25 - 20 \times 10}{25 \times 29} = {\cdot}76$$

This ·76 is the reliability coefficient of the test. We should note it carefully that both these above formulas (A) and (B) provide an estimate of internal consistency of the test rather than an estimate of the dependability of test scores. Therefore, we cannot compare the method of 'rational equivalence' with the rest three methods of determining the reliability of the test scores.

Factors influencing the reliability of tests scores

There are many psychological factors (as memory and practice effect, fluctuation of interest and attention, shifts in emotional attitude) and environmental influences (as noise, error in scoring, dissimilar conditions etc.) which affect the reliability of the test scores. These influences are varying and hence are put in a category of 'chance errors'. These 'chance errors' may cause the score to get a higher or lower value. Another type of errors are "constant errors'. The 'constant error' does not affect the reliability of the test scores. In this case all of the scores on a retest are equally affected. Say if a subject's score is 3 points more in a retest it will not mean that the self correlation of the test is affected. The correlation between the pretest and retest will remain the same, but we may say that all the scores in the second test are in error by 3 points.

For ordinary purposes we can accept 0·7 or 0·8 as a reliability coefficient. But for purposes as differentiating the individuals in group the reliability coefficient must approximate to 0·9.

The influence upon reliability of lengthening or repeating the test

We know that the average from 20 scores will be more reliable than the average from 5 scores. The greater will be the number of determinations the greater will be the reliability of the mean. On the same principles the reliability of the test depends. The greater will be the length of the test (*i.e.* the more the number of items it contains) the greater will be its reliability. We can obtain the effect of lengthening or repeating a test from the 'Spearman-Brown Prophecy Formula'.

$$r_{nn}=\frac{nr_{11}}{1+(n-1)\,r_{11}} \qquad \text{(C)}$$

(*Spearman-Brown formula for estimating the correlation between n forms of a test, and n other similar forms*)

r_{nn} = the correlation between n forms of a test and n parallel forms.

r_{11} = The reliability coefficient.

The subscripts ("11") show that the correlation is between two forms of the same test.

The use of the above formula can be illustrated as below :

A group of 200 twelve year boys were examined in a test whose self correlation is ·6. What will be the effect upon the reliability of that test if it be lengthened four times ?

Here $r_{11}=\cdot 6$
and $n=4$

$$\therefore\ r_{nn}=\frac{nr_{11}}{1+(n-1)r_{11}}=\frac{4\times\cdot 6}{1+3\times\cdot 6}=\cdot 86.$$

Thus we see if the same test be lengthened four times the reliability coefficient increases by (·86—·6) = ·26.

Instead of lengthening this test four times we can use the four parallel forms of the test and then obtain the mean of the four scores of each boy in the parallel forms. The reliability of the test obtained from these mean scores will be almost the same as we obtain by lengthening the test four times. The formula (C) can easily be used to find how many parallel forms be given or how many times the test be repeated to obtain the desired reliability coefficient.

Say as in above case the reliability coefficient of the test is ·6. Now if we want that the reliability coefficient of the test may increase upto ·86 we will have to repeat it. How many times should it be repeated ?

Here $r_{11} = \cdot 6$

$r_{nn} = \cdot 86$

$n = ?$

Applying the formula (C)

$$r_{nn} = \frac{n\, r_{11}}{1+(n-1)\, r_{11}}$$

$$\cdot 86 = \frac{n \times \cdot 6}{1+(n-1)\cdot 6}$$

or $$\cdot 86 + \cdot 516\,(n-1) = \cdot 6n$$

$$\cdot 86 + \cdot 516n - \cdot 516 = \cdot 6n$$

$$0\cdot 84n = \cdot 344$$

$$n = 4 \quad \text{(approximately)}$$

Hence the test should be repeated 4 times or 4 parallel forms of the test should be given to raise its reliability coefficient upto ·86.

The Spearman's formula can be used when the items included for lengthening the test or repeating it may be of the same difficulty, cover the same course and be almost identical to original ones. The reliability coefficient can be raised to the maximum if we repeat the test several times or lengthen it several times. But if the test items are too many in number or if it be repeated many number of times, the examinee will feel bored. He cannot attend the whole test with equal interest and thus the loss of incentive will affect the results. We can repeat our test five or six times or lengthen it six times but not more.

The Spearman's formula can also be used in ratings and judgments. Say if the two judges measure the reliability of any two attributes, we can use the Spearman's prophecy formula to estimate the reliability that can be obtained had there been ten judges or so.

The reliability coefficient from one application of a test

As we have discussed before that when no alternative form of a test is available, a single test is often splitted into two equivalent halves. For instance, the scores on odd—and on even—numbered questions or items may be totalled separately and then inter-correlated. Since the reliability of a test depends on the length of the test, hence the correlation between one half and the other half is unduly low and is usually corrected by Spearman-Brown formula :

$$r_{11}=\frac{2r_{\frac{1}{2}\frac{1}{1}}}{1+r_{\frac{1}{2}\frac{1}{1}}} \qquad \text{(D)}$$

(Spearman-Brown formula for estimating reliability from two comparable halves of a test).

Where r_{11}=the reliability coefficient of the whole test.

$r_{\frac{1}{2}\frac{1}{1}}$=the reliability coefficient of one-half of the test found experimentally.

Say if the reliability of one half of test ($r_{\frac{1}{2}\frac{1}{1}}$ is ·7 we can obtain the reliability of the whole test by using the formula (D)

$$r_{11}=\frac{2\times\cdot7}{1+\cdot7}=\cdot82$$

That means the reliability of the whole test will be ·82 if the reliability of one-half of it is ·7.

Formula (C) is same as formula (D) but in place of n in formula (C) we have 2 in formula (D) because the test is doubled.

The Index of reliability

When the same test or different parallel forms of the same test are administered to an examinee under similar conditions the mean of all these scores gives his 'true score'. The correlation between obtained scores on a given test and true scores in the function measured by the test is given by the formula :

$$r_1\infty = \sqrt{r_{11}} \qquad \text{(E)}$$

Where

r_{11}=The reliability coefficient of the given test.

$r_1\infty$=The correlation between obtained and true scores.

The symbol ∞ means that the test has been administered infinite number of times. In the above formula $r_1\infty$ represents the 'index of reliability'. It measures the correspondence between the obtained scores and the true scores. It indicates the maximum correlation which the test is capable of yielding.

Say if the self correlation of a test is ·49 then its index of reliability will be $\sqrt{\cdot 49}$ or ·7 *i.e.* the test is capable of a correlation as high as ·7. Again suppose the self-correlation of a test is ·16 then its index of reliability will be $\sqrt{\cdot 16}$ or ·4 *i.e.* the maximum correlation that the test is capable of yielding is ·4. When the index of reliability is less than ·5 we conclude that it needs improvement either by lengthening it or by repeating it.

The standard error of an obtained score

The formula for the standard error of an obtained score is :

$$\sigma_1 \infty = \sigma_1 \sqrt{1 - r_{11}} \qquad \text{(F)}$$

Where

$\sigma_1 \infty$ = The standard error of an obtained score.

σ_1 = The standard deviation of the test scores.

r_{11} = The reliability coefficient of the test.

Knowing the reliability coefficient of the test it is possible to calculate the reliability, or the limits of variation, of obtained scores from their true counter parts.

Suppose the reliability coefficient of the achievement test in English reading administered to 400 eleven year boys is ·8 and the SD of this distribution is 12. Applying the formula (F) :

$$\sigma_1 \infty = 12\sqrt{1 - \cdot 8} \qquad = 12 \times \cdot 44 = 5\cdot 28.$$

If the score of an examinee is 60 we can say with confidence (the chances are ·95) that his score actually lies between $(60 \pm 1\cdot 96 \times 5\cdot 28)$ or 49·66 and 70·34.

The dependence of the reliability upon the range and the variability of the group

We cannot compare directly the reliability coefficient of a test administered to a small group with the reliability coefficient of the same test administered to a large group. If the group is variable, it will affect the self correlation of the test. It is possible to obtain the self correlation of a test in an increased range if the self correlation of the same test in narrow range is known. The formula used is :

$$\frac{\sigma_s}{\sigma_l} = \frac{\sqrt{1 - r_{ll}}}{\sqrt{1 - r_{ss}}} \qquad \text{(G)}$$

Where

σ_s and σ_l are the σ's of the test scores in the small and large range respectively.

r_{ll} and r_{ss} = The reliability coefficient in the large and small ranges.

Suppose $r_{ss}=\cdot 6$ and $\sigma_{ll}=8$ for the fifth grade and $\sigma_l=16$ for the larger group (say children from grades three to seven).

We are to find out the reliability coefficient of the test in the wider range.

$$\frac{\sigma_s}{\sigma_l}=\frac{\sqrt{1-r_{ll}}}{\sqrt{1-r_{ss}}}$$

$$\therefore \quad \frac{8}{16}=\frac{\sqrt{1-r_{ll}}}{\sqrt{1-\cdot 6}}$$

$$\frac{1}{4}=\frac{1-r_{ll}}{\cdot 4}$$

$$r_{ll}=\cdot 9.$$

This means that the reliability coefficient of the test which is ·6 for narrow range is ·9 for a wider range (the wider range is three times the narrow range).

The test must be valid

Validity is the most important characteristic of a good examination, for unless a test is valid it serves no useful function. The validity of an examination depends on the efficiency with which it measure what it attempts to measure. A test must, therefore, accomplish the purpose the user has in mind in order to satisfy this fundamental criterion for all testing. It is not necessary that a test which is reliable may also be valid. Garret[1] has illustrated the difference between reliability and validity with the help of an example of a clock. Suppose a clock is set forward twenty minutes. If the clock is a good timepiece, the time it tells will be reliable (*i.e.* constant), but it will not be valid as judged by 'standard time'. Suppose we construct a test in mathematics for measuring the ability of the boys to apply certain principles of mathematics in solving problems in their daily living. If the actual test measured only the students, ability to recall and write down certain facts on paper, it would not be valid. Similarly a test designed to measure what a student has learned in an English course should measure his achievement in that course and nothing else. If the test is so constructed that a highly intelligent student can determine the correct answer without knowing the subject-matter, the test measures the general intelligence rather than achievement in English. Such a test is not valid. It does not measure what it is supposed to measure.

---

[1] H. E. Garrett : *Statistics in Psychology and Education*. Longman Green & Co., New York.

There are three basic approaches to validity—logical validity, empirical validity, and curricular validity. We shall discuss them one by one.

Logical validity. It is also termed as psychological validity. There are certain subjects in which it appears to be impossible to secure an objective or statistical basis of validation. In general, these subjects are in the complex fields made up of many interrelated abilities as contrasted with those practical skill areas where the tested performance either is an exact representation of, or a very similar substitute for, the instructional outcome sought. Analysis both of the desired outcome and of the prosposed test by psychological and logical methods may well reveal a sufficient degree of commonality or of similarity to justify the belief that the test constitutes a valid measure of the outcome. Such methods are followed quite frequently in such complex fields as language and the reading—study skill areas.

Empirical validity. If a test correlates with some other variable of known validity, we can say that in part it measures the same thing. Empirical validity is studied by comparing test results with a 'criterion' known to measure some characteristic of importance. For example if we wish to know whether a test will identify successful salesmen, we might give it to new employees of a wholesale hardware concern and after six months correlate the test scores with the total amount of hardware each man has sold. The criterion, 'amount sold in six months', is an index of success in selling. The size of the correlation indicates how well the test predicts this criterion. If the test does not predict it is not a measure of 'probable success in selling hardware' in his situation, and is invalid for selecting salesman.

In the case of any school subject test the teachers marks may be taken as the criterion. By correlating the test marks and the teachers marks empirical validity can be determined. But teachers marks are often unsatisfactory as a criterion because they may be influenced by many other factors than those which determine scores on the test. Such marks are influenced partly by achievement and also by attendance, class-room behaviour, recitation work, habits, general ability, sociometric status, sex and other aspects of pupils which may or may not be related to achievement test scores.

The determination of the empirical relevance of a test may include the differences in average test scores obtained by any two or more groups known to be widely separated according to ages or grades. The writer applied a general science test to the tenth class, first year science class and second year science class students of various schools. The following table shows their respective average ages, average test scores and standard deviations.

	Average Age	Mean achievement score in general science	Standard deviation	No. of pupils
Tenth class pupils	14 years 10 months	21·3	10·2	99
First year science pupils	16 years 4 months	30	9·8	108
Second year science pupils	17 years 4 months	40·3	11·04	160

The following table show that the mean achievement score of the second year science pupils is significantly higher than that of the tenth class pupils and the first year science pupils. The mean achievements score of the XI year science pupils is significantly higher than that of the tenth class pupils.

	Difference in mean achievement score	Standard error of the difference between means	Critical ratio	Remarks
XII Vs. X	·84	1·345	·63	Significant difference in favour of XII
XII Vs. XI	10·3	1·258	8·01	Significant difference in favour of XII
XI Vs. X	8·7	1·714	5·07	Significant difference in favour of XI

Thus it is clear that the mean achievement score of the pupils at the higher grades is significantly greater than the mean achievement scores of the pupils at the lower grades. This is an illustration of a valid test.

Curricular validity. A teacher who carefully and thoughtfully selects a standardized test or constructs an informal objective examinations or any other evaluation instrument for his class is attempting to insure curricular validity by making certain that the test deals with the types of educational outcomes he wishes to measure and is at the proper level of difficulty for his pupils. There are various sources of evidence to guide the teacher in considering test validity from the curricular standpoint. Among these are textbooks, courses of study, reports of national committees, and the writings of subjects and test specialists. The idea in each case is that analysis of these source materials will furnish evidence concerning the thinking of qualified educators on questions dealing with course objectives and that such an analysis affords an important objective basis for determining the outcomes to test.

The test must be objective

Objectivity is an important factor that affects both the validity and reliability of an examination. There are two aspects of objectivity which should be considered in constructing a test. The first is concerned with scoring the test. The second pertains to the interpretation of individual test items by the persons taking the test. Let us consider each of these separately.

The personal judgment of the individual who corrects the test should not be a factor affecting the score. Several people should be able to score the test and get the same results. After the key has been made and specifications have been laid down, there should be no question as to whether an item is right or wrong or partly right or partly wrong. Essay-type examinations have hardly any objectivity in making. They do not have a common objective basis for marking.

So far we were talking about the objectivity concerned with scoring or making the tests. The second aspect, equally important but more subtle and therefore more frequently neglected, has to do with student's interpretation of the items in the test. Well-constructed test items should lead themselves to one and only one interpretation by students who know the material involved. That is, a given test item should mean essentially the same thing to students who know the point in question. This is a goal that is difficult to attain. Students are prone to read into test items meanings that were never intended by the test maker. Inspite of how good a test item may appear at the start, there may be some students (good as well as poor) who interpret it wrongly. Obviously this will affect the validity. The test constructor must keep in mind that there should be no ambiguity in any test item. The items should not involve any subjective element.

So far we have discussed reliability, validity and objectivity separately. While each has been treated separately, it should be obvious by now that they are closely related and interdependent. There is no fine line of demarcation.

The test must discriminate

A test discriminates when it is constructed in such a manner that it detects or measures small differences in achievement or intelligence etc., when it picks out the good from the poor. Each item of the test should discriminate between good and poor students. In a discriminating test each item will be missed more frequently by poor students. If the good students are just as likely to miss an item as the poor students, the item does not measure in a positive direction. It is often found that poor students will answer correctly certain items that the best students miss consistently. Such items discriminate negatively. An examination of these items usually reveals that they are ambiguous or technically weak in other respects. They are not objective from the student's point of view and, therefore, do not discriminate positively. An item analysis technique is used to determine the discriminative value of each item. This technique has been discussed in detail in some later chapter.

The test must be comprehensive

The careful test maker never assumes that the instrument he has constructed is capable of measuring all of the factual knowledges or skills that a pupil has acquired in a school course. A test is a sample of certain portions of the total behaviour which the examiner considers vital to pupils' mastery in the field. A natural question then is how long any test therefore be ? How much of a sampling should it include ? In other words, how comprehensive should it be ? A test is comprehensive to the degree to which it samples sufficiently widely into the subject that the resulting scores are representative of total performance in the areas measured. In other words, a test should be comprehensive enough to be valid. It should include enough points so that it measures what it is supposed to measure. This is an easy statement to make but some what more difficult to put into practice. There is no specific formula which indicates when a test meets the criterion of comprehensiveness—it is a measure of judgment. For the class-room teacher the best practice is a careful consideration of and answer to this question. Is this test comprehensive enough to measure accurately and well what I expect it to measure ?

The test must be readily administered and scored.

Consideration must be given to the features of the test which make it readily administered and scored. It should be so designed that a minimum of student's time will be consumed in answering each item. The test item should also be constructed in such a manner that they can be scored quickly and efficiently. It is desirable that tests be subject to accurate scoring by clerical workers or other persons not conversent with their content. A convenient form of answer key or stencil should be provided for standardized test, and the manual of directions should carry complete instructions for scoring the test. The scoring keys should be arranged to that easy and accurate scoring of the tests can be accomplished. Properly spaced answers on scoring keys for informal objective examinations can be prepared by filling in the correct answers on a copy of the test and converting it into a set of strip keys cut out stencil, or a combination of the two, according to the nature of the test parts.

CHAPTER 4

ESSAY-TYPE EXAMINATION Vs. NEW-TYPE TESTS

THE TRADITIONAL OR ESSAY EXAMINATION CONTINUES TO OCCUPY an important place among the testing techniques used by the classroom teacher, although during the past few years it has lost the dominant position it occupied at the turn of the century. Scepticism concerning the traditional examination arose more than decade before 1900. Edgewor published in England during 1890 what was perhaps the first critical study of the essay test. It remained, however, for Starch and Elliott to bring the issue sharply to the front in America in 1912 by a report of marks assigned to an English examination paper by various teachers and to follow it shortly by similar reports on two other subject fields. Although it is probable that educations for various reasons somewhat misinterpreted the findings of these and many subsequent studies of the traditional examination, the fact remains that the studies very effectively called attention to a major weakness of the testing technique. In India the examination unit of the Directorate of Extension Programme for Secondary Education is responsible to bring about necessary reform in the essay-type examination. Work is still continued in this direction.

Functions of essay-type examinations

The main functions of ordinary written or essay-type examinations are as follows :

(1) They are used to rank the students tested in order of their total achievement in a given field of subject-matter.

(2) Examinations are conducted in order to discover the difficulties which the class is having so that proper attention be given to remove them.

(3) They are helpful in assessing the effectiveness of particular teaching procedures or particular methods of selecting and organizing course materials.

(4) These examinations arouse an incentive for stimulating pupils to work and also keeping teachers upto the mark.

(5) Examinations are helpful to teachers for guidance. At the time of scoring the answer-books the teacher can see where he could not make his topic clear. Then he can apply a better method for that. He can also diagnose the weak points of students and can give due attention to simplify their difficulties.

(6) Examinations require a certain amount of perseverance and industry, some docility to discipline and calmness under pressure.

These qualities become the part of the character of good boys.

Limitations of essay-type examination

1. **Rote memorization.** Essay-type examination encourages children to cram the books to pass the examination. Thus rote memory is developed. An unhealthy spirit of competition is developed among children. They try to concentrate only on those topics which are important from examination point of view. There remains no learning for doing but learning only to pass. The written examinations limit the speed of thinking of the examinee to his speed of his writing. This checking on thinking and reasoning is also due to its emphasis on mechanical memorization. It produces dislike with the school work and thus discourages the desire for further education. The next complaint is examinations dominate the whole curriculum. Though only a few per cent children go for college education after passing the High School examination and yet the course in which they are examined are the same. No special course is meant for those who are not to continue their studies after primary school education or secondary school education. It is often seen that the boys sit idle for the whole year and begin their studies too seriously a few weeks before the examinations with the result that they spoil their health and sometimes turn maladjusted.

2. **Unreliability of written examinations.** When the two examiners are asked to mark the same answer-book, there is every possibility that their marks differ to a large extent.

Again if we construct two tests, including the same course, to examine a particular child it is likely that he will obtain different marks on two different occasions.

Elles[1] had 61 teachers score an examination consisting of four essay questions in geography and history, and eleven weeks later had them score the same answers again. Reliability coefficients, obtained by correlating the first and second scores assigned by the same teacher, aged from 0·25 to 0·51 for the four essay questions. This and other evidence showing wide differences in the two sets of scores assigned by the same person, led him to conclude that the same individuals vary from time to time to their judgments about as widely as different individuals vary.

This happens because our tests are not reliable. The main causes of unreliability are :

(*a*) *Examinee's mental and physical state of mind.* The examinee's mental and physical state of mind affects his examination performances. If two tests of the same difficulty be given to the same child on two consecutive days, he may score the first day better than on the second day. Many external and irrevelant happenings may temporarily depress his work on the second day.

(*b*) *Inadequate sampling.* The examination or test paper may not include the questions covering the wide range of the examinee's knowledge. We can call it as inadequate sampling of examinees knowledge. Suppose a written examination paper contains ten questions to cover the course. It is certain that these ten questions will not be sufficient to examine the achievement of the child who has covered the whole course thoroughly. A poor child may also guess these questions and cram them. Both the children will get the same marks. Where is then the discrimination which is one of the main objectives of the examinations ? If on the next day another ten questions be set on the same paper or if on the same day that paper may not contain those expected questions, there is every chance that poor child may fail. For a reliable test number of questions should be greater.

---

[1] Walter C. Elles : *Reliability of Repeated Grading of Essay-Type Questions,* Journal of Educational Psychology, 21, pp. 48-52, January (1930).

(*c*) *Divergent scale of marking.* The standard of marking of different examiners may vary. Two different examiners may give different marks for the same answer. Unless the marks given by the examiners be uniformly standardized, it is not possible to make the test reliable.

(*d*) *The personal element.* Some examiners may give credit for accuracy of facts, better organization and formulation of thoughts, others may appreciate a good handwriting, neat and tidy work. Naturally if the same answer-book is given to these two types of examiners they will give different marks on it. Again if the examinee has mentioned those theories which the examiner likes most there is possibility of his getting higher marks. These are all the personal elements which make the test least reliable.

Starch and Elliott, who had 142 teachers score identical copies of an English examination paper, found that the scores based on 100 per cent for perfection ranged from a low of 50 to a high of 98. In another study, they found that 115 teachers rated a geometry paper from a low of 28 to a high of 92.[2]

3. Lack of validity in essay-type tests. A test in arithmetic, translated from English language to Hindi language, was given to a certain class of children. They could not get the better scores because the language was complicated. Next time the same test in a simplified Hindi was given to the same class. This time children could solve it in a better way. No doubt the test was reliable but was not valid in the former case. First time the test was a rather stiff vocabulary and reading test as well as artithmetic test. So it was not a valid measure of arithmetical ability of children. A valid test is that which claims to test that ability for which it is meant. It has already been pointed out in the first chapter that a test is used to measure the students outcome resulted from our teaching. We teach with a certain objectives and thus try to change the behaviour of the students. Tests are helpful to indicate whether those objectives have been achieved or not. A test may be called valid if it measures those objectives for which it is meant. Unfortunately most of the essay tests are not constructed in a scientific way with the result that they measure something else from what is proposed to measure. Hence essay tests are not valid.

---

[2] Daniel Starch and Edward C. Elliott : *Reliability of Grading High School Work in English and Mathematics*, School Review, September 1912 and April 1916.

4. **Bluffing factor.** Pupils who do not know answers to essay questions are prone to respond in terms calculated to cover up their lack of information if not actually to mislead the teacher. Such responses, which tend to vary in plausibility directly in relation to the intelligence of the bluffer, may take the form of discussion concerning content closely related to that covered by the question, of very incomplete answers which by repetitious statements and copious illustration may give a sense of completeness, and various other devices. Whether bluffing is or is not desirable is not the issue. Certainly bluffing is resorted to in great or small degree by all persons on some, if not many occasions. To the extent, however, that bluffing is actually successful on essay tests, the examination results are less accurate measures of pupil achievement.

Advantages of the essay tests

Only the major and rather commonly accepted advantages of the essay tests will be discussed here.

1. **Ease of construction and administration.** Essay tests are commonly considered easy for teachers to prepare and to administer. Teachers typically give minimum of time to the preparation of essay questions. Sometimes the questions are not even formulated until immediately before the examination is to be given. Little or no time is taken for telling pupils how to take the essay test. However, essay tests prepared and administered with a minimum of effort are likely to have such resulting disadvantages that the saving of time and labour may well be at the expense of testing efficiency.

2. **Measurement of higher mental abilities.** Advocates of essay type tests insist that the discussion-type questions have values not possessed by the informal objective test—in that they call for comparison, for interpretation of facts, for criticism, for defence of opinion, and for other types of

higher mental activity. The purpose of the written test is primarily to ascertain whether the student has accurate knowledge and a considerable amount of understanding about a variety of matters in terms of their interacting relationships but not basically to determine whether he knows certain highly detailed facts and whether he has not routine course requirements. It is only through essay tests we can measure the expression and accurate knowledge of facts of the child. The freedom of response that the essay test question allows is considered to be one of its fundamental characteristic. By the nature of the question the student is required to survey his own background of related information and to select the related facts and organize them for expression in his own words. It is important, however, that the freedom of selection, organization, and expression be suited to the measurable outcomes of the course.

3. Adaptability to school subjects. It is possible to use the essay examination for practically all subjects of the school curriculum, for the question and answer method is widely acceptable. Some types of outcomes, such as arithmetic skills, handwriting skills, reading ability and others, cannot be stated directly by this device, but the factual backgrounds for them frequently can be so tested. As a matter of fact, the essay test procedure is often used in scoring arithmetic examples by the use of arbitrary decisions in scoring for correctness of the result or correctness of the method, in given partial credit for answers not entirely correct, and for various other ways.

Suggestions for improving the essay tests

(1) The purpose of the examinations must be understood by both the teacher and the pupil. The emphasis of the essay examination should be definitely on thought, reasoning and other types of mental activity as applied to the materials of the course.

(2) The content of the examination should be governed by its purpose. It should measure those objectives and students outcomes for which it is meant.

(3) The preparation and selection of suitable essay-type questions should consume at least as much time as is required to score the answers.

(4) The questions should be relevant and demand short responses. It is only then the scoring can be done objectively.

(5) Definite rules should be formulated that will as far as possible control irrelevant factors in scoring the papers. The careful use of scoring rules will bring about a definite decrease in the inaccuracy of the pupil scores.

NEW-TYPE TESTS

We may now be convinced that essay-type examinations have many limitations. They can hardly form a good test unless they are modified in the light of the suggestions already pointed out in the previous section. New-type or objective-type test can be used with advantage to supplement the essay-type examination. There is a dilemma whether essay-type examinations be given a prominent position or the new-type tests be given the priority. A practical 'modus vivendi' can be achieved by compromise. Neither we can discard essay-type tests altogether nor the new-type tests. Each must be supplemented by the other. Before discussing this point any more let us see the various types of New-type tests or the objective tests that are commonly used. It will also be our aim to discuss the advantages and disadvantages of these types of tests.

New-type tests or objective tests

The principal types of objective test items used by class-room teachers may be listed as follows :

(1) **Recognition types**
    (a) Multiple-choice
    (b) Alternative-response
    (c) Matching
    (d) Classification

(2) **Recall types**
    (a) Simple recall
    (b) Completion

We will discuss each of these types of tests separately.

Multiple-choice items. A multiple-choice item consists of question or incomplete statement followed by several possible answers. The student must select the best or correct answer in accordance with the direction given. Multiple-choice items are of four kinds:

(1) One right answer
(2) One best answer
(3) Analogy-type
(4) Reverse multiple-choice

(1) One right answer. This is the simplest kind of multiple-choice item. The student is required to identify the one correct response listed among several that are totally wrong but not obviously wrong.

Illustration

Directions. *Each of the items mentioned below is followed by five answers. Select the best possible answer in each case and put its letter in the bracket on the left of the item. Item (1) is solved for you as an example.*

(1) The apparatus used for measuring the atmospheric pressure is called a :

(a) Thermometer (b) Barometer
(b) (c) Pyrometer (d) Lactometer
(e) Hydrometer

(2) The weight of one cu. ft. of water is 62·5 lbs. and one cu. ft. of cork is 15 lbs. The specific gravity (relative density) of the cork is :

(a) 4·17 (b) ·24
( ) (c) ·417 (d) 2·4
(e) 15

(3) The weight of a piece of iron in air is 10 lbs. and its apparent weight in water is 8·7 lbs. Its specific gravity is :

(a) ·13 (b) 1·87
( ) (c) 7·7 (d) 8·7
(e) 10

(2) One best answer. The variation requires the student to select the best response from a series of several responses. It is perhaps the most valuable of the several forms of multiple choice items. It has often been criticised, however, because the selection of the best response involves judgment. This is not a valid criticism. The fact that it does require judgment or inferential reasoning or complete understanding to select the best response is the most important feature of the item.

Illustration

Directions. *Each of the following items is followed by five answers. Select the best possible answer in each case and put its letter in the bracket on the left of the item.*

(1) The mouth of a vacuum flask is closed with a cork because :

(a) Cork is cheap.
(b) Cork is durable.
( ) (c) Cork is washable.
(d) Cork is a bad conductor of heat.
(e) Cork is lighter in weight than other substances.

(2) In cold climates woollen clothes are used because :

(a) They are durable.
(b) They give heat to the body.
( ) (c) They diminish the loss of heat from the body.
(d) They are heavier than cotton clothes.
(e) They are good looking.

(3) Analogy-type items. The analogy item is an adaptation of a mathematical type of abbreviation (4 : 2 :: 10 : ?). The student is required to discover the relationship that exist between the first two parts of the item. He then applies this relationship to the third and fourth parts. The third part is given, but the fourth must be selected from the several choices on the basis of relationship existing between the first two parts. This type of item has been used more often in intelligence testing than in achievement tests.

Illustration

Directions. *In the following item determine the relationship between the first two parts of the item. Then apply this relationship to the third and fourth parts by selecting the proper choice (A, B, C, D, or E) to go with the third part. Place the letter of your choice in the bracket at the left.*

(1) Hack saw : saw blade : : jack plane : ?

(a) Chisel
(b) Plane iron
( ) (c) Cap screw

(d) Block plane
(e) Jointer plane

(4) Reverse multiple-choice item. The reverse multiple-choice item differs from other forms in that the student is required to choose the poorest or wrong response from among several correct responses. This variation is especially adapted for that type of material in which there are several correct or best responses.

Illustration.

Directions. *Each of the following items is followed by five answers. Select the right answer in each case and put its letter in the bracket on the left of the item.*

(1) It is not true that the image formed by a convex mirror is always :
    (a) virtual
    (b) diminished
( ) (c) erect
    (d) the same distance behind the mirror as the object is in front
    (e) not at the same distance behind the mirror as the object is in front

(2) Which of the following statements is not true of carbon dioxide ?
    (a) Carbon dioxide is a colourless gas.
    (b) Carbon dioxide has a very faint smell.
([illegible]) (c) Carbon dioxide has a very slight taste.
    (d) Carbon dioxide can be turned into a liquid at ordinary room temperature by the application of pressure.
    (e) Carbon dioxide is lighter than air.

(3) Which of the following is not a method of producing carbon dioxide ?
    (a) Burning carbon in air.
    (b) The action of dilute acids on all carbonates.
( ) (c) The heating of sodium.
    (d) The breathing of animals carbonate.
    (e) Fermentation.

(4) Which of the following is not true of acids ?
    (a) They turn blue litmus red.
    (b) They are sour to the taste.
( ) (c) They are corrosive.

(d) They react chemically with carbonates.

(e) They all contain oxygen.

Advantages of multiple-choice items

(1) Multiple-choice item can be used to measure what one can recognize, which represents a much wide field than what one can recall.

(2) The guessing factor does not present as much of a problem in multiple-choice items as in several other types. If the number of alternatives or choices are four or less than four a correction formula eliminates the guessing.

The formula is :

$$S=R-\frac{W}{N-1}$$

Where S = Score corrected for guessing.
R = Number of right responses.
W = Number of wrong responses.
N = Number of alternatives or choices.

Let us suppose that there are 20 multiple-choice items in a test. Each item has four alternatives or choices from which the student is required to select the correct one. Let us further suppose that a student answers only 10 items correctly. If we give one mark for each correct item then of course he will secure 10 marks in our test. It is just possible that he may have guessed some of the items and thus by chance got them correct. To eliminate this chance factor or guess factor we apply the correction formula.

In the present case :

$$R=10$$
$$W=10$$
$$N=4$$

And therefore $$S=10-\frac{10}{4-1}$$

$$=10-\frac{10}{3}=\frac{20}{3}=7.$$

Three marks have been deducted for guessing. The larger is the number of alternatives the lesser is the guessing and hence we need not apply correction formula for those items which have five or more than five alternatives.

(3) In the case of multiple-choice items scoring can be made entirely objective. It is well adapted to machine scoring.

(4) The multiple-choice item can be designed to measure effectively the students ability to interpret, discriminate, select and make application of things learned. It can be used to measure understanding, judgment and inferential reasoning ability. For these purposes it is generally the most valuable of the several types of objective test items.

Limitations of multiple-choice items

(1) Many test makers do not know or do not take the time to utilize multiple-choice items properly. In too many instances the item is constructed to measure memorization only, rather than application.

(2) It is often difficult to construct a good item so that one and only one response is the correct one.

(3) The multiple-choice test is space consuming and time consuming.

(4) It is not completely free from guessing.

Alternate response or true-false items

The true-false item consists of a declaratory statement or a situation that either true or false (right or wrong). The student decides which of the two possible choices is the correct one and places his answer accordingly. There are two main kinds of true-false items.

(1) Regular true-false.

(2) Modified true-false.

(1) Regular true-false. This is a simple form of true-false item. It usually consists of a simple statement that may be either true or false. The student is required to indicate whether or not the statement is true. Let us assume that the teacher has taught the class about 'solution'. If he wants to know whether the pupils have followed his lesson or not and whether the principles and facts are clear to them or not, he can apply the true-false test. The teacher can hand over the printed copies of the test to the pupils or to avoid the labour he can dictate the items. The test should have at least 40 items. When using a true-false test for an examination on which the results of one year's work is to be decided, there should be at least 100 items to eliminate the unreliability arising from mere chance and guess work. The teacher will ask the pupils to put a plus sign at the left of the statement if it is right and zero (0) if the statement is wrong. The whole test including 40 items can be finished in 15 minutes. The necessary directions should be given by the teacher before starting the test. After the test is completed, a lively discussion can be held.

Illustration

Directions. *About one-half of the following statements are true and about one half are wrong. Mark each true statement with a plus sign* (+) *and each false statement with zero* (*0*) *at the left. Do not mark the statements that you do not know. Ask no question. Do not guess. Item* (*1*) *is solved for you as an example.*

(+) (1) Water is the best solvent of all the liquids.
( ) (2) Wax is soluble in water.
( ) (3) That thing which dissolves in the solvent is called the solute.
( ) (4) There can be a solution of a gas in the liquid.
( ) (5) Copper sulphate is soluble in water.
( ) (6) Sand is soluble in Benzene.
( ) (7) etc.

Such 40 items can be put in this test.

As evident from the test there are two possibilities of guess in this case *i.e.* right or wrong. Suppose out of these 40 items, a pupil gets 30 items right and 10 items wrong, he will secure only 20 marks. The examiner infers that each wrong response means that two guesses have been made and deducts two marks to make up for the one got right by guess work.

Total marks $=40$
Wrong responses $=10$
Marks secured for 30 right responses $=40-10-10=20$,
The formula applied for scoring is :

$$S=R-\frac{W}{N-1}$$

where S = Score corrected for guessing.
R = Number of right responses.
W = Number of wrong responses.

N=Number of alternatives (as in this case there are two alternatives, right or wrong.)

Applying this formula the pupil who gets 30 right responses and 10 wrong responses will secure 20 marks.

$$S=30-\frac{10}{2-1}, \text{ because } N=2$$

$$=30-10=20.$$

Let us assume that a pupil realizes that he knows practically nothing about the subject-matter. Maximum that he can guess is fifty per cent *i.e.* he can get 20 answers right and 20 wrong. It does not mean that he will secure 20 marks. According to the above formula his marks will be :

$$S=20-\frac{20}{2-1}=20-20=0.$$

(2) Modified true-false. Modified true-false items may be designed to require the student to mark the items that are true but to modify the false items by (1) crossing out the word that makes the statement false or (2) identifying the word that makes the statement false and listing another word that would make it true.

Illustration

Directions. *Some of the following statements are true and some are false. If the statement is true, underline the 'T' at the left and do no more. If the statement is false underline the 'F' and do two more things.*

1. In blank 'A' insert the word that makes the statement false.

2. In blank 'B' insert the word that would make it true. The first item is answered as an example.

1. T F In large cities newspapers are printed on cylinder machines.

(A) (Cylinder) (B) (Rotary)

2. T F Plants take in oxygen and give out hydrogen during the night.

(A)________ (B)________

Advantages of true-false items

1. The true-false item is easy to construct and hence extensively used.

2. It can be scored readily in an objective manner.
3. It is especially useful when there are only two choices or alternatives governing a particular point.
4. It can be used to sample a wide range of subject-matter. True-false items can be answered in a short time and hence large number of such items can be included in test.
5. Modified true false items require the students to exercise judgment and utilize his knowledge in situations that call for understanding rather than mere memorization.

Limitations of true-false items

1. Plain true-false item encourages guessing.
2. It hardly measures the true achievement of a student.
3. A true-false test is likely to be low in reliability unless it includes a large number of items.
4. It is difficult to construct if the material is in any way controversial.

Matching items. As the name implies, this type of item requires the matching of two or more sets of material in accordance with given directions. The common matching item consists of two columns of words or phrases The student is required to match each item in one list with the item in the other list to which it is most closely related.

Illustration

Directions. *In the following sections there are two columns. For each statement or question in the first column there is an answer in the second column. Find the answer and place its letter in the brackets given in front of the item in the first column. The first items in section I is done for you as an example. Answer as many questions as you can in each section in the same way.*

Section I. What is the chemical formula for one molecule of :

(L)	(1) Nitric oxide	(A)	$H_2 SO_4$
( )	(2) Ammonium chloride.	(B)	$KMnO_4$
( )	(3) Nitrous oxide.	(C)	$NaCl$
( )	(4) Nitric acid	(D)	$CaCO_3$
( )	(5) Calcium hydroxide	(E)	$(NH_4)_2SO_4$
( )	(6) Potassium permanganate	(F)	HNO_3
( )	(7) Sulphuric acid .	(G)	NH_4OH
( )	(8) Hydrogen sulphide (or sulphurated hydrogen)	(H)	NH_4Cl
( )	(9) Ammonium sulphate	(I)	H_2S

( ) (10) Sodium chloride

(J) N_2O
(K) $Ca(OH)_2$
(L) NO

Section II. What units are the following measured in :

( ) (1) Density
( ) (2) Power of a lens
( ) (3) Latent heat
( ) (4) Work
( ) (5) Quantity of heat
( ) (6) Potential difference
( ) (7) Force
( ) (8) Power
( ) (9) Pressure
( ) (10) Electric current

(A) ft. lb.
(B) cm. per sec. per sec.
(C) ampere
(D) volt
(E) dyne
(F) foot pound per sec.
(G) grams per c.c.
(H) calories
(I) dioptre
(J) calories per gram.
(K) pound weight
(L) pound weight per square inch

Advantages of matching items

1. The matching item can be made totally objective and can be scored quickly. Matching items form a reliable and valid test.

2. When it is properly constructed, the guessing factor is practically eliminated.

3. It is relatively easy to construct. A large number of responses can be included in a small space and with one group of directions.

4. It can be used, with advantages, to measure the students' ability to recognize relationships and make associations and for naming and identifying things learned.

Limitations of matching items

1. Since the phrases or clauses must necessarily be short, the matching exercise provides a poor measure of complete understandings and interpretation.

2. It is inferior to the multiple choice item in measuring judgment and application of things taught.

3. It is likely to over-emphasize the memorization of facts.

Classification item

Classification item is a variation of the matching item.

Illustration

Directions. *In the following items underline the word that does not belong to that category. First item is solved as an example.*

1. Apple, Mango, Cat, Orange, Grapes.
2. Horse, Cow, Ox, Dog, Mountain.
3. Lake, Stone, River, Ocean, Pond.

Recall items

In the case of recall item the student supplies his answer directly. Sometimes the students may be asked to answer the question directly in one word or a group of specific words. This is called a simple recall item. Sometimes the students may be asked to supply a word or words missing from a sentence. This is called a completion item. Care needs to be taken in preparing such tests. The blanks should be such that only one and one word may fit in. Sometimes the first letter of the omitted word may be given. Below are given examples both of simple recall and completion items.

Simple recall items

Directions. *Read the following questions and put their answers in the space after the questions. The first item is solved for you as an example.*

1. Which type of balance is used for measuring the mass of a body ?.........physical balance.................
2. What temperature would 10°C read if converted to Fahrenheit scale ?.........................
3. At what temperature would the reading of a Fahrenheit and a Centigrade thermometer be the same ?...........
4. What is the liquid part of the blood called ?................
5. Which method of pollination is most suitable in the case of the rose flower ?.....................................
6. A group of about fifty eggs was found. The eggs were yellow in colour, bluntly conical in shape, prettily marked with zig-zag stripes. Name the insect or the animal whose eggs they could be..................................

Completion items

Directions. *In each of the following items certain blanks marked as (A) or (B) are to be filled in. You are to put your* *answers in the brackets on the left side. First item is solved for you as an example.*

1. Plants take in.........A......and give out.........B......... during the night.
   (Oxygen) A
   (Carbon dioxide) B
2. When a solid is dissolved in a liquid the solid is called the.........A.........
   ( ) A
3. The weight of one c.c. of water is one gram at......A......° C.
   ( ) A
4. The solubility of a substance in a liquid is defined as the weight in grams of the substance which will dissolve in...A...grams of the liquid to form a saturated solution at a given temperature.
   ( ) A
5. The poles of a simple magnet are of......A......strength.
   ( ) A
6. The oxides which behave either as acidic or as basic oxides are called......A......oxides.
   ( ) A
7. The cutting teeth are known as......A......
   ( ) A
8. The teeth whose function is tearing flesh are known as ....A.........
   ( ) A
9. Fish taken in.........A.........for respiration from the water by means of their......B......
   ( ) A
   ( ) B
10. When carbonates of metals are decomposed by heat ......A......is given off and an,..B...of the metal is left.
    ( ) A
    ( ) B

Advantages of recall-type item

1. The recall-type item can be used to measure the retention of specific point. It demands accurate information.

2. It is comparatively easy to construct.

3. Recall items can be substituted for recognition items when it is desired to have to the student recall out right the information being tested.

4. Students are generally familiar with recall item and as such they find it easy to score them.

5. These items have high discriminating value. The guessing factor is minimized in well prepared items.

6. Recall items form a reliable, valid and objective test.

Limitations of recall-type item

1. Unless care is exercised in constructing the recall items, the scoring is apt to become subjective.

2. It is difficult to measure complete understanding with the simple types (simple recall and completion).

3. The student may know the material being tested but have difficulty in recalling the exact word needed to fill in a certain blank.

4. Simple recall item tends to over-emphasize verbal facility and the memorization facts.

CHAPTER 5

A PROGRAMME FOR CONSTRUCTING AND STANDARDIZING ACHIEVEMENT TEST IN ANY SCHOOL SUBJECT

What is an achievement test

AN ACHIEVEMENT TEST IS AN INSTRUMENT DESIGNED TO MEASURE relative accomplishment in a specified area of work. It can be classified as :

(1) General achievement test.
(2) Diagnostic test.

In this chapter we will discuss the steps that are involved in constructing the general achievement test in any subject area. In section (A) of this chapter broad outline for constructing the test have been given. The writer[1] has constructed a 'general science test' for "a comparative study of the standard of general science in the secondary schools of England and India (U. P.) for pupils of a chosen age range." A brief account of this research study will be helpful to the readers for constructing and standardizing the tests in other school subjects. An account of this study has been given in section (B) of this chapter.

SECTION A

The theory of constructing and standardizing the general achievement test

Before discussing the construction of the general achievement test in any school subject let us be familiar with this term. "A general achievement test is one, designed to express in terms of a single score a pupil's relative achievement in a given field of achievement."[2]

---

[1] D. S. Rawat : *Unpublished M. A. Education Thesis*, Birmingham University, (1957).

[2] H. E. Hawkes, E. F. Lindquist and C. R. Mann : *The Construction and Use of Achievement Examination*, Houghton Miffin Co, Boston.

The general achievement tests are used to rank the pupils in order of their achievement in the field of school subject. Their achievement is not measured in terms of absolute units. Again an arbitrary standard such as a 'passing grade' is not used to measure the achievement. The main flaw in the school examinations is that they measure the achievement of the pupils in terms of passing grade. If a pupil secures 33% marks he passes otherwise he fails. The test may not be reliable. It is possible that if the same test be repeated next time that particular pupil may pass. So with this much of discussion we conclude that the students should be classified and ranked according to their comparative achievement This principle is adopted in the case of general achievement test.

In the definition of the general achievement test the word 'field of achievement' is used. Let us clarify what it means. We can prepare a test to cover a broad course on general science or in branch of it, say 'physics' or even we can select one topic from that branch, say, 'heat'. But in all three cases the test will be general achievement test if it covers the broad course in that specific field. The characteristic of a general achievement test, then, as it is here considered, is that performance in a given 'field' whatever the limits of that field, is expressed in terms of a single score.

Before constructing the general achievement test in any school subject we ask two questions to ourselves.

1. What to measure ?
2. How to measure ?

What to measure ?

The first problem, *viz*., 'what to measure' leads us to the main objectives of instruction in that specific field. In the first chapter it has already been pointed out that we teach any subject with certain objectives in view so that we may be able to bring about the expected change in the behaviour of the students whom we teach. In chemistry the objectives which the instructors are commonly trying to reach include teaching students to acquire a fund of important facts and principles ; to understand the technical terms commonly appearing in chemical publications ; to be able to apply important chemical principles to appropriate situations ; to express chemical reactions by means of equations involving chemical symbols and formulae ; to be skilful in certain laboratory techniques. Any adequate examination programme for chemistry will provide means for discovering how far each of these objectives is being attained. What is true for chemistry is true for all school subjects. After all our main function for applying the test is to see how far the students have been able to achieve the objectives.

So our first step in constructing the general achievement test in any school subject will be :

1. To define the objectives of instruction in that specified field.

2. To examine the course content. We may call it curriculum analysis. An analysis of curriculum and text books is necessary to see whether these objectives fit in the pattern of our course of study.

3. To analyze and define each objective in terms of expected student outcomes. It requires the listing of the elements that are a part of each objective. We define each element in terms of students behaviour. We are also to list items of subject-matter related to each element.

How to measure ?

This involves as how to construct a measuring instrument that will best do the measuring. For constructing the test we proceed in the following way. These steps are discussed in brief here because a detailed discussion of these steps are given in section (B) of this chapter.

1. **Construction of individual items.** Keeping in view the objectives, behaviour patterns and course contents objective items are framed. These items should be valid. They should not be ambiguous. It is always better to put the items in the completion form in the pilot test because the various expected responses of the students can be achieved and these responses can be helpful in putting the items for the final test in multiple-choice form or matching form. If the test has different types of items they should be put in different sections. For the pilot test or the first try-out test there should be sufficient number of items (not less than 100).

2. **Clear and concise directions for each type of questions should be given.** Instructions for taking the test should be

given on the cover page. The print of the test should be bold and distinct. There should be no printing mistakes or omissions.

3. The pilot test is applied in a sample of wide range of variability. The time limit for scoring test should be so adjusted that at least 80 per cent pupils have time to consider all the items of the test.

4. The tests are scored. Correction for chance formula is applied wherever necessary.

5. The reliability and the validity of the test is determined. The various methods for determining the reliability and the validity of the tests have already been discussed in the previous chapter.

6. Each item of the test is analyzed to find out its discriminating value. Items having negative discriminating value or no discriminating value are discarded. This technique has been discussed in a later chapter.

7. The final test is prepared in the light of the item analysis. Only those items are retained in the final test which discriminate positively. It is also kept in view that there are sufficient items to measure each objective. Some items can be modified or converted from one type to another, say, completion type items can be put in the form of matching type or multiple-choice. Directions and instructions can also be modified in the light of the pilot test. Time limit for the test should be written on the cover page. Time limit for the final test can be fixed with the help of the time taken by the testees for the pilot test.

8. The final test is again applied to another sample of wide range of variability. This time the sample should be large enough (not less than 2000).

9. The tests are scored. The reliability and the validity of the test are calculated.

10. The average of the students' scores is determined. This is called a norm. The norm is used to compare the scores of different examinees Norms may be school-grade norms, age norms, sex norms and urban area or rural area norms. If we calculate the average score of the twelve year old children in a certain test we will get age norm. If we calculate the average score of eighth class students in a certain test we will get grade

norm. Similarly sex norms, urban area norms and rural area norms can be calculated. A test is called a standard test only when its norms are ready because with the help of the norms of the test we are able to label the achievement of any particular student in that test. So norm is a form of standard unit for that test and hence the test having a norm is called a standardized test.

11. After calculating the norms interpretations are made. Interpretations vary according to the nature and purpose of the tests. Readers are referred to section (B) of this chapter to have an idea of interpretations on the basis of test results.

Sometimes the examiner or the test constructor tries to find the 'Educational quotients' and the 'Achievement or Attainment quotient' of the students. These terms have been defined, below :

Educational quotient and achievement or attainment quotient

Before defining these terms we need to know another term *viz.* 'Educational age'.

Educational age. The average of any number of attainment (or subject) ages provides an educational age, which represents an objective assessment of the average level reached in a number of subjects as measured by standardized tests.

The achievement age of any student in a particular subject is shown by the actual level of his chronological age. If a child of eight years passes an achievement test in algebra meant for a child of ten years, his achievement or attainment age in algebra is ten years If, on the other hand, his level of achievement in that subject is the same as that of a normal eight year, his achievement age is only eight years.

The educational age of a child is obtained by adding together his achievement ages in various subjects and dividing the total by the number of subjects.

$$\text{Educational age} = \frac{\text{total number of achievement ages in various subjects}}{\text{number of subjects}}.$$

But for the record cards and for class-room purposes we should calculate the attainment age or achievement age in separate subjects separately and thus calculate the educational age for each subject separately. It does not make sense to average two Arithmetic ages, one English reading age and three Geography ages. This is grossly unfair to pupils who had a marked weakness either in Arithmetic or in English reading or in Geography and still he attains the normal educational age.

The Educational quotient is another measure.

$$E.\ Q = \frac{E.\ A.}{C.\ A.} \times 100$$

where

E. Q. = Educational Quotient.
E. A. = Educational Age.
C. A. = Chronological Age.

So educational quotient is the ratio derived from comparing the educational age (E. A.) with the chronological age (C. A.). The multiplication by one-hundred is done to get rid of the decimal point.

Achievement quotient (A. Q.) is the ratio derived from comparing the educational age (E. A.) with the mental age (M. A.), so that,

$$A.\ Q = \frac{E.\ A.}{M.\ A.} \times 100.$$

The multiplication by one-hundred is done for the same reasons. For the clear understanding of chronological age and mental age, the readers may see the chapter 'intelligence tests' of this book.

Both these measures (E. Q. and A. Q.) have limitations. In making a comparison of educational age with chronological age to obtain an educational quotient we are assuming that chronological age is correlated more closely with school progress than it really is. Thus dull children always have low educational quotients ; and the same measure can mark the unrealized margin of talent in bright children. Similarly the achievement quotient, though better for certain purposes of assessing the levels of which the pupil is capable, gives peculiar results with some pupils. Bright childrens are usually expected to do better than dull children but with the use of achievement quotient dull senior pupils often have a better chance of achieving normal or even advanced A. Q.'s because one compares attainment only with mental age and hence gives them the benefit of experience. The achievement quotient is really unsatisfactory if we compare pupils of unequal ages. For example, if we compare the older dull pupil of C. A. 11, M. A. 8, E. A. 8½, who easily gains an A. Q. of 106 ; with the younger brighter pupil of C. A. 8, M. A. 11 and E. A. 9, who has only an A. Q. of 82 ; we find that we have set on extremely difficult task to the brighter younger pupil to bring his A. Q. upto 100. Progress in school is dependent on maturity and years of schooling, and in the cases quoted above an advantage of three years of these influences is given to the older backward pupil, while a handicap of three years enters into the calculation of the A. Q. of the younger pupil. For the younger brighter pupil we have considered only his advanced intelligence but this could not help him if he had not been taught certain school work.

SECTION B

In this section an actual working procedure for constructing and standardizing a general science test has been discussed. It is beyond the scope of this book to outline the project in full. Only those steps have been discussed which are relevant here.

Purpose of the project

When we study the history of the teaching of general science we are forced inevitably to look to the west. Although important discoveries in several fields, such as medicine, astronomy, agriculture, metallurgy and mathematics were made in ancient India, yet science did not develop beyond a certain stage in our country and languished. It is generally felt that there is a great need for bringing about the improvement in general science if it is to survive as a school subject and if the country's need for technicians and scientists is to be met. This need has been felt timely by the All India Council for Secondary Education (now taken a shape of the National Council of Educational Research and Training) which has taken seriously the question of improvement of science teaching in Indian schools. Not only this, but teachers from our country have gone to U. K., U. S. and Canada to study the standards of general science in the secondary schools, with a view to get help from them, positively, in teaching and learning of general science in the secondary schools.

The investigator has also tried to make an experimental contribution in the same direction with this comparative study of the standards in general science of the pupils of a certain age group in some secondary schools in India and England—*A comparative study of the standard of general science in the secondary schools of England and India U. P. for pupils of a chosen age range.*

The average chronological age of an Indian pupil reading in the X class of the secondary schools in U. P. (India) is 14 years 10 months and the average chronological age of an English pupil reading in the fourth form of the secondary schools is 14 years 11 months.[1] These two average chronological ages are almost equal and, therefore, 14+ has been taken as constant. The investigator made an effort to compare the achievement in general science of the X class pupils (U P.) and the fourth form pupils in the secondary schools of England. It was also his aim to find out particular class grade in the Higher Secondary Schools or the Intermediate Colleges in U. P. (India) that comes to the general science achievement level of the fourth form of the secondary schools in England.

The best procedure for carrying out the project could be to construct a general science test suitable for both the fourth form pupils in England and the pupils in the Higher Secondary Schools in U. P. (India) and this plan was adopted.

The investigation, though important, is not without difficulties. At a particular age (here it is 14+), there are wide variations in the amount of teaching given, syllabi and standards of teaching in the secondary school of England and India (U. P.). This accounts for validating the test for both the countries. The second of these difficulties could be eliminated by constructing a general science test, keeping in view the syllabi of Indian secondary schools as well as those of England, with reference to the courses they have in the X class and the fourth form respectively. When topics are common for these two groups of pupils, naturally any difference in the general science achievement of these can only be interpreted in terms of the differences in the standard of teaching or the amount of the teaching given to them.

---

[1] See the later part of this section.

would be to measure the increment of desirable activities that he can and does perform as a result of this study. However, only a few attempts to devise tests for such a purpose have so far been made.

Analysing and defining each objective in terms of expected student outcomes

All the objectives or the outcomes cannot be measured adequately by using one single test. The present test is designed to measure only the following objectives :

(1) Acquisition of a knowledge of the principles and facts of courses.

(2) An understanding of important technical terminology.

(3) The ability to identify structures and processes and their functions.

(4) Ability to draw reasonable generalizations from experimental data new to students.

(5) The ability to apply scientific principles to situations new to student.

All these objectives can be analysed in terms of behaviours or outcomes that individuals are expected to attain and in terms of the content by which the behaviours are to be developed. An analysis of these objectives in terms of behaviour pattern has been done below. Having defined the expected outcomes that individuals are expected to attain, it has been decided what content will measure the particular behaviour pattern. Some examples of the contents in the form of test items have been given below :

(1) *Acquisition of a knowledge of the principles and facts of courses.* (a) It may involve the ability to distinguish between the facts and misconceptions in the material which the pupil reads. The following are some of the items which are used in the present test to measure this ability.

Which of the following substances supplies the greatest amount of protein ?

(a) rice (b) white of egg
(c) vegetable (d) fruits
(e) milk.

Pilot test— Second part } Item No. (37)

Animals become rickety and have bad teeth if there is a deficiency of :

(a) vitamin D (b) vitamin A
(c) vitamin E (d) vitamin B
(e) vitamin C

Pilot test—
Second part } Item No. (36)

Pilot test—Part one—Section VIII.
Pilot test—Part two—Item No. 14, 19, 20, 21, 28, 32, 33, 34, 35, 36, 37, 39, 48, 50, 51.

(b) The acquisition of information may include a test of pupil's ability to state desired information in response to direct questions. It was kept in view that the questions be couched in phraseology which is not a verbatim reproduction of text books. Exercises of this sort can be in the 'essay questions' form but in so doing the test becomes less objective because different examiners of the pupils' answers often give very different evaluations to responses. To minimize this weakness the exercises have been put in the 'completion form'. The following are the examples of this sort of exercises.

Pilot test—Part three.

Directions. *In each of the following items certain blanks marked as (A) or (B) are to be filled in. You are to put your answers in the brackets on the left hand side.*

Item No. (4)—The solubility of a substance in a liquid is defined as the weight in grams of the substance which will dissolve in............A............grams of the liquid to form a saturated solution at a given temperature.

( ) A

Item No. 2, 3, 5.

(2) *An understanding of important teachnical terminology and symbols.* (a) It may involve the pupil's ability to recall a term when confronted with its definition or description. Examples of exercises which give evidence of the understanding of technical terms are given below :

Pilot test—Part one—Section III.

Directions. *On the left hand side there are ten processes.*

On the right hand side there are ten descriptions not in the same order. In the brackets provided write the letter which indicates the statement corresponding to each process in turn.

( ) (1) conduction

( ) (2) electrolysis

( ) (3) filtration.

( ) (4) convection

( ) (5) refraction

( ) (6) distillation

( ) (7) radiation

( ) (8) evaporation

( ) (9) plasmolysis

( ) (10) osmosis

(A) The method used for separating an insoluble substance from a liquid.

(B) The bending of a ray of light which travels obliquely from one transparent medium to another.

(C) The way in which heat is carried from one point to another in solids.

(D) The method used for separating a dissolved substance from a liquid.

(E) The main way in which heat is carried from one point to another in a liquid.

(F) The collapse and contraction of the protoplasm of a cell due to the passing of water out of the cell.

(G) The process by which liquids and substances in solution tend to pass through permeable and semi-permeable membrane.

(H) The process of splitting up a substance chemically by means of an electric current.

(I) A process used for obtaining pure water from a solution.

(J) The way in which heat reaches us from any source without appreciably warming the intervening medium.

Pilot test—Part one—Section IV.

In the next exercise the term which best represents the definition or description is to be checked.

Pilot test—Part two—Item No. (45).

The process by which water is lost from the cells of a plant by evaporation into the intercellular spaces, and the escape of the vapour through the stomata is known as :

( ) (a) respiration (b) osmosis
(c) imbibition (d) plasmolysis
(e) transpiration.

In the following exercises the pupil is to recall the omitted term.

Pilot test—Part three.

Directions. *In each of the following items certain blanks marked as (A) or (B) are to be filled in. You are to put your answers in the brackets on the left hand side.*

Item No. (6).

The oxides which behave either as acidic or as basic oxides are called.........A.........oxides. ( ) A.

Pilot test—Part 3—item No. (7), (8), (9).

The following items which are in the simple recall form also give evidence of the understanding of technical terms.

Pilot test—Part four.

Directions. *Read the following questions and put their answers in the space after the questions.*

Item (4).

What is the liquid part of the blood called ?.........

Pilot test—Part 4—Item No. (7).

In chemistry courses, students are expected to associate symbols with the chemical elements and compounds which they represent. This ability has been tested in the present test by asking the pupils to give the correct symbols for an element or compound.

Following are the examples.

Pilot test—Part two.

Item No. (24).

Which of the following represents four molecules of oxygen correctly ?

(a) $4O_2$
(b) $2O_2$
(c) $4O$
(d) O_2+O_2
(e) O_3+O

Item No. (25).

Which of the following is the correct formula for copper nitrate ?

( ) (a) $Cu_2 (NO_3)$
(b) $Cu NO_3$
(c) $Cu_2 (NO_2)$
(d) $Cu (NO_3)_2$
(e) $Cu (NO_2)$

Pilot test—Part one—Section (I).

Pilot test—Part two—Item No. (23).

Pupils are also expected to balance the equations which are given in the incomplete form.

Examples are given below :

Pilot test—Part three.

Directions. *Complete and balance the following equations* :

(11) $H_2 SO_4 + \ldots\ldots\ldots\ldots = Na_2 SO_4 + \ldots\ldots$
(12) $H_2O + Na = \ldots\ldots\ldots + H_2$
etc.

It is also desired to test pupil's ability to understand the units. For testing this ability matching exercise has been constructed.

Pilot test—Part one—Section (II).

What units are the following measured in :

( ) (1) density
( ) (2) power of a lens.
( ) (3) latent heat
( ) (4) work
( ) (5) quantity of heat
( ) (6) potential difference
( ) (7) force
( ) (8) power
( ) (9) pressure
( ) (10) electric current

(A) ft. lb.
(B) cm. per. sec. per. sec.
(C) ampere.
(D) volt.
(E) dyne
(F) foot pound per c.c.
(G) grams per c. c.
(H) calories.
(I) dioptre
(J) calories per gram.
(K) pound weight.
(L) pound weight per square inch.

The provision of more units than here are words is helpful to eliminate successful guessing answering the questions.

(3) *The ability to identify structures and processes and their functions*. In biology course units are expected to identify plant and animal forms. This requires the photographic reproductions of plants and animals. In the present test this could not be done. The pupils may be expected to be familiar with plant and animal structures and processes. In physics courses, pupils are expected to be familiar with machines or apparatuses and to identify their parts and the functions of those parts. In chemistry courses, students are expected to be acquainted with various chemical processes, with the different structures involved in those processes, and with the functions of the structures.

In the present test the diagrams of the plant and animal structures and of the apparatuses have been drawn. The pupils are asked to write the letter of the pointer in the brackets in front of the items.

Examples. Pilot test—Part one—Section V, section VII.
Pilot test—Part three—Item (17).

The following items have been included in the test to measure the ability of the pupils to identify the functions of the parts of animal and plant structure.

Examples. Pilot test—Part two.

Item (26).

Bile is secreted by the cells of the :

( ) (a) duodenum (b) rectum
(c) liver (d) stomach
(e) ileum.

Item (38).

Blood is carried away from the heart by vessels known as :

(a) veins (b) capillaries
(c) arteries (d) trachea
(e) nerves.

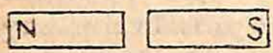
Pilot test—Part one—section VI.

Pilot test—Part two—Item No. (27), (29), (40).

(4) *Ability to draw reasonable generalizations from experimental data new to students.* This ability can be tested by presenting the experimental data and asking the pupils to draw reasonable generalizations from it. The following items have been included in the test to measure pupil's ability to draw reasonable generalization from experimental data new to them.

Examples. Pilot test—Part two.

Item No. (44).



A bar magnet was cut into two pieces as shown in the figure. Which of the following statements is true ?

( ) (a) first piece will have only N pole left. (b) second piece will have only S pole left.

(c) the magent is destroyed (d) both pieces will act as magnets.

(e) there will be no poles in either piece.

Pilot test—Part four.

Item No. (6).

A group of about fifty eggs was found. The eggs were yellow in colour, bluntly conical in shape, prettily marked with zig-zag stripes. Name the insect or the animal whose eggs they could be.

Pilot test—Part two—Item No. (16), (18), (43).

(5) *The ability to apply scientific principles to situations new to students.* This objective involves the application of past experience in solving new problems. Evidence of the ability to apply scientific generalizations may be obtained by presenting the pupil with a new problem situation and asking him to state what will happen or why it will happen. Sometimes the answers to such questions involves long statements and hence scoring can hardly be done objectively. To minimize this weakness for each problem five answers have been given and the pupil asked to pick up the right one.

Examples. Pilot test—Part two.

Item No. (8).

In cold climates woollen clothes are used because :

( ) (a) they are durable (b) they give heat to the body.

(c) they diminish the loss of heat from the body. (d) they are heavier than cotton clothes.

(e) they are good looking.

Item No. (13).

Sometimes we treat water with chlorine gas because :

( ) (a) it makes water colourless. (b) it helps the floating particles on the water to settle down.

(c) it removes the bad smell of water. (d) it kills harmful germs.

(e) it adds taste to the water.

Pilot test—Part two.

Item No. (6), (7), (22), (30), (31), (41), (42), (46), (47), (49).

Some items to test this ability, have been put in the completion form and simple recall form.

Example. Pilot test—Part three.

Directions. *In the following item two blanks marked as (A) and (B) are to be filled in. You are to put your answers in the brackets on the left hand side.*

Item No. (10).

( ) A When carbonates of metals are decomposed by heat.................A...............is given
( ) B off and an..............B.........of the metal is left.

Item No. (18) to (22).

Example. Pilot test—Part four.

Item No. (5).

( ) Which method of pollination is most suitable in the case of rose flower ?

Item No. (8).

Some numerical questions have also been put under this head in the present test. The pupil is able to solve the numerical questions in science if he understands the application of the principles clearly. A pupil who does not understand the application of the principle of Archimedes will not be able to solve the following numerical question.

Pilot test—Part two.

Item (3).

The weight of a piece of iron in air is 10 pounds and its apparent weight in water is 8·7 lbs. Its specific gravity is—

( ) (a) ·13 (b) 1·87 (c) 7·7 (d) 8·7 (e) 10.

Examples. Pilot test—Part two.

Item No. (2), (4), (5), (9), (10), (11), (12), (15), (17).

Pilot test—Part four.

Item No. (2), (3).

Analysis of the curriculums and text books

The chief purpose of the test was to study the comparative achievement of the pupils of the fourth form in the secondary schools in England and of the pupils reading in the high school classes and intermediate classes in India. The testing programme was confined only to certain technical and grammer schools in England and to the high schools and intermediate colleges existing in Uttar Pradesh (India).

To make the test meaningful it was necessary to analyse the curriculum and text books taught in these schools. In England both the curriculum and text books vary from school to school. The teachers possess a great degree of freedom regarding the curriculum. "No freedom that teachers in this country possess is so important as that of determining the curriculum and methods of teaching. Neither the Minister nor the local Education Authority exercise authority over the curriculum of any school beyond that of agreeing the general educational character of the school and its place in the local educational system. The governers in consultation with the head of the school, are normally responsible for the general direction of the conduct and curriculum of the school, but in practice the curriculum is settled by the head in co-operation with the assistant staff."[2]

Although head teacher in England are free to decide the curriculum, it is markedly similar in character and content in schools of the same type. There are various factors which account for this similarity in the character and content of the curriculum of various schools of the same type. The views of the parents carry much weight in England. Spens Report says[3] "An English School which departs widely from average practice can do so only with the support of a body of parents of unusual views". Social climate, political philosophy and public opinion also carry a good weight in the determination of the curriculum.

---

[2] W. O. Lester Smith: *Education*, Penguin Books, (1957), pp. 161-62.

In England much good has been done by various associations which are constituted to promote study of particular aspect of the curriculum—Science, History, Mathematics, English, etc. These associations are helpful to teachers for the exchange of ideas and thus resulting a common faith in certain values.

Another important influence in determining the curriculum is the Inspectorate. Inspectors are appointed by Her Majesty in Council. "They do not and are not expected to, promulgate in school any particular form of curriculum or any particular methods at the behest of the Minister ; they report to the Minister what they find and think, not what they suppose he wishes them to find or think ; their reports cannot be altered without their consent (though the Minister can decline to issue them if he wishes to do so). The corollary of this is that from time to time inspectors hold different opinions and express different views in schools. Some measure of diversity is harmless ; in the long run, indeed, it is essential to progress. What would be harmful would be any suggestion that an inspector had the right to press his views on a school and require it to carry out its work in his way. Schools do well to remember that an inspector's advice is really advice, and advice that need not be followed."[4]

An Inspector's contact with schools are of two main kinds. Periodically schools are inspected with meticulous thoroughness, and on such occasions the inspector visits as one of a team, carefully chosen to include experts in those branches of knowledge represented in the school's curriculum. But no less important, and often more valuable are the informal visits that inspectors make. Every school is allocated to some one inspector, who keeps in touch with it and, from time to time, arrange if it seems likely to be helpful for specialist or other inspectors to visit and give the teachers concerned the benefit of their expert knowledge.

---

[3] *Report of the Consultative Committee on Secondary Education*, H. M. S. O.

[4] *Education in 1949* (H. M. S. O.), pp. 96 and 94.

There is yet another kind of influence that helps to shape curriculum, educational policy and practice, namely that of various consultative and advisory bodies.

The curriculum in the secondary schools in India does not vary from school to school in the same state because it is prescribed by the Board of Education of that state. Teachers have hardly any voice in deciding the curriculum for a particular grade. Whereas curriculum construction should be the duty of teachers who know the needs of pupils but in India curriculum is decided by those who are surely gifted by nature but have hardly any experience of class room teaching. Thus in the secondary schools of a particular state in India there is uniform and rigid pattern of curriculum.

It has already been pointed out that the average chronological age of the fourth form pupils in the secondary schools in England is equal to the average chronological age of the tenth class pupils in the secondary schools in India. Therefore those common topics were selected, for the construction of the test, which are taught both in the fourth form in England and in the tenth class in India. While constructing the items based on these common topics, the question papers for the home examinations in various forms were kept in view.

Some science tests were also helpful to give an idea of the suitability of the items for this particular standard. Some science text books were also consulted to construct the appropriate items.[1] The curriculum analysis has been outlined in the original thesis. Because of the lack of space it has been omitted here.

---

[5] (a) Gupta, R. K. : *An attempt to construct and standardise test in general science with reference to some of the aims of teaching the subject at ordinary level, G. C. E.*, M. A. Thesis, London University. England.

(b) Mealings, R. J. : *General Science Test*, Birmingham University, England.

THE FIRST TRY-OUT

Administration of the test

The test for the first try-out contains 155 items including physics, chemistry and biology. To have a sample of wide range of variability three schools of different types were selected for carrying out the test.

Saltley Grammar School is a mixed school of high all round academic attainments. King's Norton Grammar School is a boys' school where the importance of science teaching has long been emphasised. St. Philip's Grammar School is a Roman Catholic boys' Grammar school drawing its pupils from a fairly wide area.

The test was administered in the month of January, 1957. It took about 12 days to test the sample of 125 boys and girls. The investigator personally administered the test but the science teachers of the classes were kept there to ensure seriousness and honest effort. The same type of directions were given to the pupils in different schools.

Before applying the test in the schools five boys of the fourth form were tested individually. The time for completing the test was recorded in each case. The average of these timings was 80 minutes and hence it was thought proper to keep the time limit of this preliminary test as eighty minutes.

For bringing about certain improvements in the test the pupils were asked to underline the words that they did not understand. They were also instructed not to guess and to omit those items which they were unable to answer. The test was set out in the omnibus form. It is believed that the pupils in the fourth form are accustomed to examinations or class room work for two periods of 40 minutes each and hence fatigue can hardly be significant in the test score.

As far as possible disturbances within or without the room that might interfere with the administering of the tests were prevented. The time limit as set in the directions for giving the test was strictly observed.

Scoring the test

The answers of this test are given in the column form. For scoring the test items quickly and efficiently a strip key was prepared. In the case of matching type, recall type and completion type, one point was given for each correct response. Correction for chance formula was applied in the case of multiple choice type of items. Chances of guessing the correct answers vary with different item forms. There is little if any chance of guessing, or at least of making a pure guess, on recall item forms and matching item forms. The chance is 50-50 on an alternate-response item and this was the reason for not putting this type of items in the present test. The chance is only one in five for a multiple choice item with five alternatives.

The correction for chance formula is

$$\text{Score} = \text{Rights} - \frac{\text{Wrongs}}{N-1} \text{ or } R - \frac{W}{N-1}$$

where N represents the number of possible answers to an item. In the present case N is five.

Correction for chance is ordinarily used with the true-false test and the multiple-choice test consisting of items that have as few as three alternatives. It need not necessarily be used with multiple-choice items having four or more alternatives, as the chance of making a correct guess is not great in such tests. But there is evidence that the validity of a standardised test tends to be raised slightly when a correction for guessing is carried out.[6] By calculating the means for corrected scores and uncorrected scores of the pupils in the present test it was seen that there was a reliable difference between the means and hence it can be concluded that the correction for guessing is necessary.

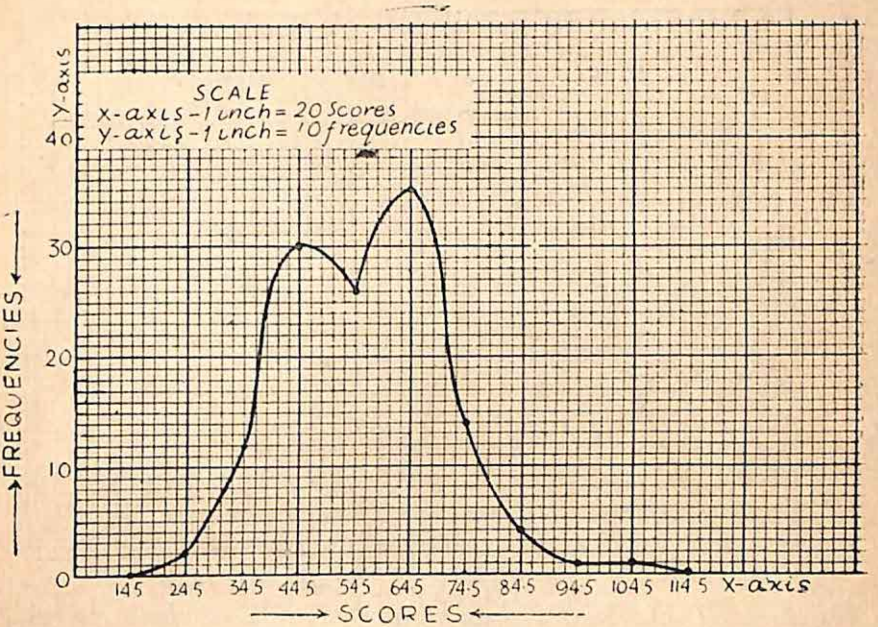
Mean, Median and Standard Deviation of the scores made by 125 pupils (boys and girls) upon an achievement test in general science (correction for chance not being applied) are as below :

Mean	$(M_1) = 56{\cdot}26$
Median	$(Md_1) = 56{\cdot}61$
Standard Deviation	$(S.D._1) = 14{\cdot}538$
Skewness	$(Sk) = -{\cdot}07$.

This skewness is almost negligible. The frequency polygon (Graph 1) of the distribution of the scores is almost a normal curve with a slight skewness on the negative end.

---

[6] *Educational Tests and Their Uses*, 'Reviews of Educational Research', 5 : 482 (1935).



Graph No. 1

Frequency Polygon of the distribution of the scores made by 125 boys and girls of fourth form in the General Science achievement test.

(Scores not being corrected for guessing)

Mean, Median and Standard Deviation of the scores made by 125 pupils (boys and girls) upon an achievement test in general science (scores being corrected for chance) are as below :

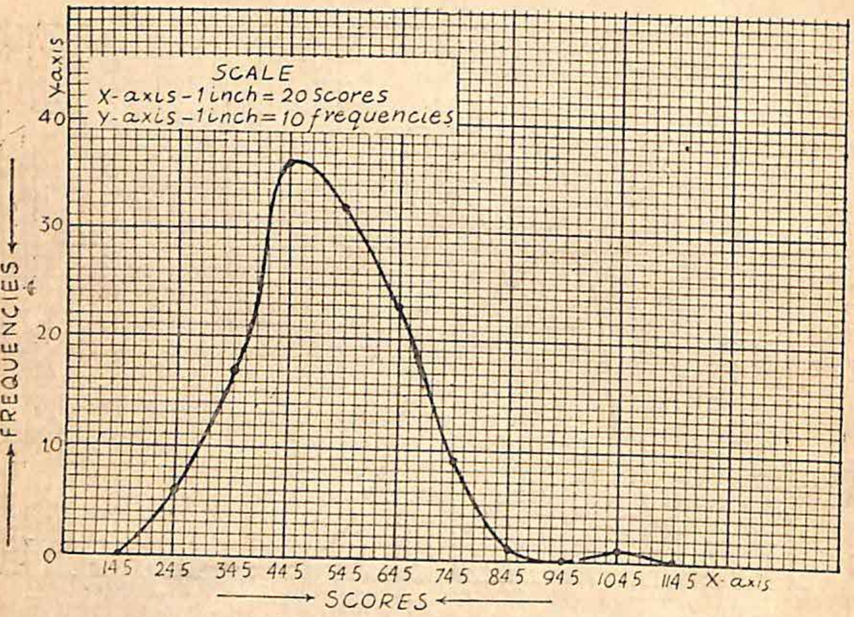
Mean $(M_2) = 51{\cdot}4$

Median $(Md_2) = 51{\cdot}59$

Standard Deviation $(S.D._2) = 13{\cdot}7$

Skewness $(Sk) = -{\cdot}05$. This skewness is almost negligible. The frequency polygon (graph 2) of this distribution is almost a normal curve with a slight skewness on the negative end.

The coefficient of correlation between the two sets of scores, *i.e.*, scores being corrected for chance and scores not being corrected for chance is ·94.



Graph No. 2

Frequency Polygon of the distribution of the score made by 125 boys and girls of the fourth form in the General Science achievement test.

(Scores being corrected for guessing)

The reliability of the difference between the means of the two sets of scores: Scores being corrected for chance and scores not being corrected for chance—

	Scores not being corrected for chance	Scores being corrected for chance
No. of pupils	125	125
Mean Score	$M_1=56\cdot26$	$M_2=51\cdot4$
Standard Deviation	$S.D._1=14\cdot4$	$S.D._2=14$
Standard Error of Mean	$S.D._{m_1}=\frac{14\cdot4}{\sqrt{125}}=1.28$	$S.D._{m_2}=\frac{14}{\sqrt{125}}=1.24$
Difference between Means	——→ 4·86	←——
Correlation between two sets of scores	——→ ·94	←——

Formula for standard error of the difference between two correlated means :

$$\begin{aligned} SE_d &= \sqrt{(S.D._{m_1})^2+(S.D._{m_2})^2+2\gamma\ S.D._{m_1}\times S.D._{m_2}} \\ &= \sqrt{1{\cdot}28^2+1{\cdot}24^2-2\times{\cdot}94\times(1{\cdot}28)\times(1{\cdot}24)} \\ &= \sqrt{3{\cdot}1760-2{\cdot}983936}=\sqrt{{\cdot}192064} \\ &= {\cdot}438 \end{aligned}$$

$$\therefore\ t \text{ ratio}=\frac{4{\cdot}86}{{\cdot}438}=11.1$$

$$df=125-1=124.$$

From the table for '*t*' values it is seen that for 124 *df* the value of '*t*' is 2·62 at the ·01 level. Our '*t*' 11·1 is far greater than 2·62 and hence is very significant. It is, therefore, concluded that by using the correction for guessing or chance there is a significant change in the mean score of the pupils.

The reliability of the pilot test

The reliability of a test depends upon the consistency with which it gauges the abilities of those to whom it has been applied. A reliable test should yield essentially the same score when administered twice to the same student, provided, of course that no learning occurs while the test is being taken the first time or no learning or forgetting takes place between testings. There are four main methods of determining the reliability of an achievement test as discussed before :

(1) Test-Retest method.
(2) Parallel forms method.
(3) The Split-half method.
(4) The method of 'Rational Equivalence'.

The investigator used the method of 'Rational Equivalence' for determining the reliability of the test.[7] This method stresses the inter-correlations of the items in the test and correlations of the items with the test as a whole and is free from the limitations which other methods possess.

---

[7] Kuder, G. F. and Richardson, M. W. : *The Theory of Estimation of Test Reliability*, Psychometrika, 1937, 2, 151—160.

The formula that is used for determining the reliability of the test is

$$\gamma = \frac{n}{n-1} \times \frac{S.D_t^2 - \Sigma xy}{S.D_t^2}$$

in which

γ = reliability coefficient of the whole test.

n = number of items in the test.

$S.D_t$ = the Standard Deviation of the test scores.

x = the proportion of the group answering a test item correctly.

$y = (1-x)$ = the proportion of the group answering a test item incorrectly.

The chart showing the proportions passing each item (x) and proportions failing each item (y).[8] was prepared. For each item, 'x' and corresponding 'y' were multiplied and their sum was determined which was called Σxy.

In this case

$$S.D.t = 14\cdot538$$

$$n = 155$$

$$\Sigma xy = 25\cdot394$$

$$\text{Reliability coefficient } \gamma = \frac{155}{154}\left[\frac{(14\cdot538)^2}{(14\cdot538)^2} - \frac{25\cdot394}{(14\cdot538)^2}\right]$$

$$= 1\cdot006\left[1 - \frac{25\cdot394}{211\cdot353}\right]$$

$$= 1\cdot006(1 - \cdot12)$$

$$= 1\cdot006(\cdot88)$$

$$= \cdot8852$$

$$\text{Index of reliability} = \gamma_{1\infty} = \sqrt{\gamma}$$

$$= \sqrt{\cdot8852}$$

$$= \cdot94.$$

No hard and fast rule can be stated to the question as to how high the reliability coefficient must be before it is acceptable. But for our purpose the 'achievement test in general science' must have a reliability coefficient of ·90.

---

[8] See the original thesis.

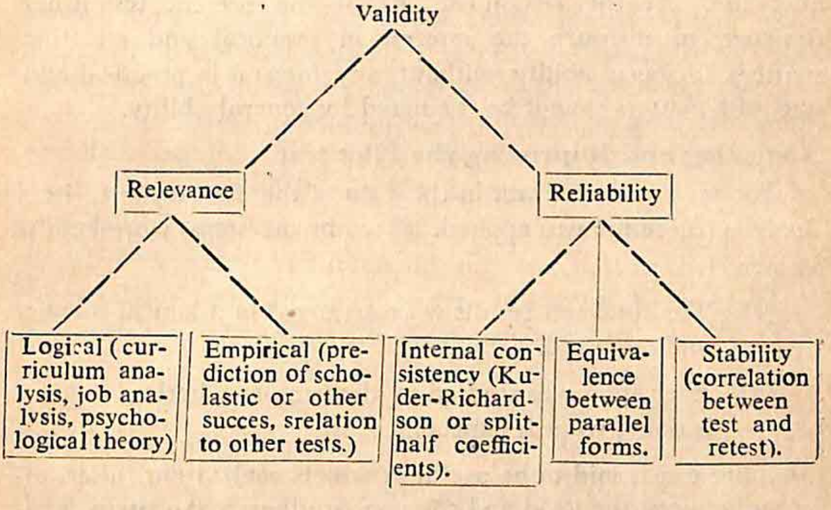
The validity of the Pilot test

The validity of a test is the degree to which it measures what it is intended to measure. This definition can be broken into two parts :

(1) "the degree to which it measures",
and (2) "what it is intended to measure".

The degree to which a test measures anything and measures it accurately, is the 'reliability' of the test. "What it is intended to measure" is the criterion for the 'relevance' of the test.

The following diagram[9] shows the components of validity and methods of estimating them.


Validity
Relevance
Reliability
Logical (curriculum analysis, job analysis, psychological theory)
Empirical (prediction of scholastic or other succes, srelation to other tests.)
Internal consistency (Kuder-Richardson or split-half coefficients).
Equivalence between parallel forms.
Stability (correlation between test and retest).

The reliability of the pilot test has already been discussed. The empirical relevance of the test has been determined in the present case. I.Qs. of the pupils has been taken as the criterion against which the predictive validity of the test has been determined. But before choosing any criterion we have to see whether there are some common elements in the criterion and the test. If the common elements are not present then the criterion will be a false index of predictive validity. In his comprehensive article[10] on 'Psychology and the Teaching of Science', Dr. Peel concludes from the studies of Ward, Ellis, Berridge, Khan, Earle and Zyve that there is no strongly delineated group factor of science ability. Pupils possessing high general ability with some interest in practical and scientific matters would seem to make the potential scientists. Taking this as a hypothesis the investigator tried to correlate the I.Q. scores of the pupils with their corresponding general science scores.

---

[9] Remmers, H. H. & Gage, N. L. : *Educational Measurement and Evaluation*, p. 123. Harper & Brothers, New York, 1954.

The I.Qs. of only 107 pupils were available and hence the product moment coefficient of correlation between the I.Qs. and the corresponding general science scores of these 107 pupils has been determined. The coefficient of correlation is ·36. This correlation is not high. It may be because the sample is not adequate. Another reason can be that the present test is not designed to measure the interest in practical and scientific matters. Science ability without any interest in practical and scientific matters cannot be predicted by general ability.

Analyzing and Improving the Pilot test

For improving the test in the light of the first try-out, item analysis technique was applied. Two broad steps were kept in view.

(1) The total test results were arranged in a logical manner and analyzed.

(2) Each item was studied to determine whether or not it is doing what is expected of it.

That test is said to be useful in which each item discriminates between the good and the poor students. An item analysis is helpful to the test constructor for the following reasons :

(*a*) The poor item can be discarded without impairing the value of the test. Thus much time and labour is saved.

(*b*) It is useful in diagnosing the nature of the individual difficulties of the students. Remedial instructions can be given accordingly.

---

[10] Peel, E. A. : *Psychology and the Teaching of Science*, Brit. J. Psy., Vol. XXV, Part III, Nov. 1955, pp. 139–141.

(*c*) Some discarded items of the former tests can be used in the latter tests where they prove to be useful.

(*d*) The test constructor can easily know which items are good for being put in the test and which to be discarded.

For analyzing the items the following procedure was adopted.

The total test score was taken as the criterion for grouping the papers. Pupils securing good scores on the test were assumed to be bright students and boys securing low scores on the test were assumed to be poor students.

The test papers were arranged in the descending order. The paper having the highest total score was placed on top. The paper with the lowest score was placed at the bottom. The test papers were grouped in terms of equal thirds for the purpose of comparison. The groups were named as upper one-third, middle one-third and lower one-third. The student item chart was prepared as given in the original thesis. The numbers across the top relate to the individual test items (there were 155 items in this test). The numbers in the left-hand column relate to the test scores of the students who took the test (125), with the highest score on top. For each student the items, which he marked correctly, were checked. A sign (√) was placed in each square for each individual except that under the items which he marked incorrectly. The purpose of this step is to record how many pupils in each group get the item right and how many either get it wrong or omit it. The individual columns and the rows of the upper and lower groups were then added separately. Now the problem was to determine statistically whether the difference between the upper and lower groups on a single item was significant. This was done by finding out the percentage difference between the two groups and dividing by the standard error of the differences. This computation was done for each of the items. The procedure is illustrated by using the results for the second item taken from the table [Appendix (d) of the thesis]. The same procedure was followed for each of the remaining items.

Let N_1=number in upper group marking item correctly.

X_1 = total number in upper group.

P_1 = percentage of upper group marking item correctly.

Q_1 = percentage of upper group marking item incorrectly.

N_2, X_2, P_2, Q_2 = same designations for lower group.

First the percentage figures for both the groups were determined.

For the second item—

$$P_1=\frac{N}{X_1}\times 100=\frac{33}{44}\times 100=75$$

$$Q_1=100-75 \qquad =25$$

$$P_2=\frac{N_2}{X_2}\times 100=\frac{20}{44}\times 100=45{\cdot}5$$

$$Q_2=100-45{\cdot}5 \qquad =54{\cdot}5.$$

The formula used for finding the standard error of the difference is

$$\text{S.D.}_{(diff.)}=\sqrt{\frac{P_1Q_1}{X_1}+\frac{P_2Q_2}{X_2}}$$

By substituting the values of P_1, Q_1, P_2, Q_2, X_1 and X_2 in the above formula

$$\text{S.D.}_{(diff)}=\sqrt{\frac{75\times 25}{44}+\frac{45{\cdot}5\times 54{\cdot}5}{44}}=9{\cdot}9$$

The percentage difference between the two groups was divided by the standard error of the difference :

$$\text{Critical Ratio (CR)}=\frac{P_1-P_2}{SD_{diff.}}=\frac{75-45{\cdot}5}{9{\cdot}9}=\frac{29{\cdot}5}{9{\cdot}9}=4{\cdot}51.$$

This resultant ratio exceeds two and hence this item is regarded as discriminating. In this way all the items were analyzed. The items having a ratio of less than two were eliminated from the test. In the table [Appendix 2(d) of the thesis] all those items marked (√) were retained for the final test.

and those items marked (×) were eliminated. Thus 84 items out of 155 were retained for the final test.

Exception was made in case of some items which had a discriminating value less than two. Item (1) and (2) in ‘Section VII’, Part I, have discriminating values less than two but these items were retained because they are discriminating in the upper group. Less than fifty per cent pupils in the upper group and only about twenty five per cent in the lower group could distinguish between the arrangements of the voltmeter and an ammeter in the circuit.

In part (2), item (35) has the discriminating value 1·8 but on inspection it seems that this is worth retaining. Seven per cent pupils in the upper group could solve it and none in the lower group. It clearly distinguishes between the very bright pupils and the rest. In part (3) the seventeenth item is in the matching form. Third item of this section was solved by almost an equal number of pupils both in the upper group and the lower group. This item is evidently an easy one, still it was not rejected because matching form of exercise should have at least five responses[11]. When there are less than five responses, the material can usually be measured more effectively by another type of items (multiple-choice, preferably) but it was not appropriate to put this exercise in the multiple choice form.

THE FINAL-TEST

By using the item analysis technique it was discovered whether each item was discriminating between good and poor pupils or not. Item analysis was also helpful to discover items that were too easy, even for the poor students or to discover if any items were too difficult, even for the best students. By using the statistical technique it was found that out of 155 items only, 84 items were worth retaining. Exercise 17th of the third part was placed in the first part. Fourth and sixth sections of the first part were completely removed from the final test. In the second part of the test twenty-five items, of the multiple-choice form, were retained. In the final test the items were arranged in a slightly different order to avoid the falling of the correct responses in a particular pattern. The good items of the third and the fourth part of the previous test were put in the completion form in the final test. These items could have been put in the simple recall type. Practically there is no difference between the completion form and the simple recall form. In the case of simple recall form the examinee is sometimes tempted to give the answer in a full sentence but in the completion form he is required to recall and supply one or more key words that have been omitted from statements and hence scoring is more objective and easy in the latter case.

---

[11] Herbert E. Hawkes, Lindquist, E. F. and Mann C. R. : *The Construction and Use of Achievement Examinations*, Boston, Houghton Mifflin Company, 1936, p. 150.

The final test items are intended to measure the following objectives :

(1) The acquisition of a knowledge of the principles and facts of courses.

(2) An understanding of the important technical terminology and symbols used in this field.

(3) The ability to identify structures and processes and their functions.

(4) The ability to draw reasonable generalizations from experimental data.

(5) The ability to apply scientific principles to situations new to the students.

The following items in the final test are retained to measure these various objectives.

First objective :—The acquisition of a knowledge of the principles and facts of courses :—

Final test—Part II—Item No.—(6), (7), (11), (13), (14), (15), (23), (25).

Second objective :—An understanding of the important technical terminology and symbols used in this field.

Final test—Part I—Section—(I), (II), (IV).
Part II—Item No.—(2), (3), (9), (20).

Third objective :—The ability to identify structures and processes and their functions.

Final test—Part I—Section—(III), (V), (VI), (VII).
Part II—Item No.—(12), (16), (17).

Fourth objective :—The ability to draw reasonable generalizations from experimental data.

Final test—Part II—Item No.—(4), (5).
Part III—Item No. (7).

Fifth objective :—The ability to apply scientific principles to situations new to the students.

Final test—Part II—Item No.—(21), (22), (10), (2), (8), (18), (19), (3), (24).
Part III—Item No.—(4), (4′), (5), (8), (8′), (8″).

Out of these 84 items which were retained in the final test, 32 items were in Physics, 29 items in Chemistry and 23 items were in Biology.

Most of the items in the final test are in the matching form and hence it was not possible to grade the items according to difficulty value. Still it was tried to avoid any significant piling up of the items at any one level of difficulty.

Adequate provision for identification of pupil on the title page was provided. Adequate directions were provided to the pupils for the entire test and for each section of it. The provisions for the pupil's responses were made as adequate and convenient as they could be, both from the point of view of the pupils in taking the test and from the point of view of scoring the test. The provisions for responses were arranged in vertical columns for convenience in scoring.

In the pilot test there were 155 items and the time-limit was eighty minutes. In the final test there were only eighty-four items and hence the time-limit was kept as forty minutes.

Administration and Scoring of the Final test in the British Schools

One purpose of administrating the test in the British schools was to study in a limited way, the comparative achievement of boys and girls, in Grammar and Technical schools, in general science. It was also aimed to study the comparative achievement of pupils in the Grammar and the Technical schools in general science.

Four Schools (1) Secondary Technical School (West Bromwich) with 53 boys and 13 girls, (2) King Edward's School (Stourbridge) with 82 boys, (3) Holly Lodge Girls' Grammar School with 77 girls and (4) Hinckley Grammar School with 50 boys and 37 girls were selected for the investigation. All these pupils were selected from the fourth form of the different schools.

The Secondary Technical School (West Bromwich) is a mixed school. King Edward's School (Stourbridge) is a grammar school in a country town. Holly Lodge Girls Grammar School is a girls' school of a good repute. Hinckley Grammar School is a big mixed school where it was possible to get a wide range of variation among the pupils in their achievement in general science. So the pupils 312 in number were expected to be a fair sample of the fourth form pupils.

The test was administered in the last week of March and the first week of April. The investigator personally administered the test in all the four schools with the help of the school teachers.

There are numerous ways in which administration errors may impair test scores. As far as possible those sources of error in the test administration were given due consideration. Probably the greatest single source of error in test administration is incorrect timing of the test. The timing error of the test was prevented by emphasizing the need for precise timing and by giving necessary direction for it on the cover page.

A second large source of error in administration is lack of clarity in the directions to the examinees. Confusion of directions is, in some respects, more serious than mistakes in timing. Unclear directions may lower the scores of either the entire group or of certain pupils in varying amounts. Since the investigator administered the test personally hence there was no doubt in the uniformity of giving the directions to the pupils. The same procedure in presenting the directions was followed.

A third source of error in the administration of objective test is failure to make clear to pupil what they are expected to do about guessing. Necessary instructions were given to the pupils not to guess. It was explained that guessing reduces the score.

A fourth source of error in test administration is variation in the physical conditions under which tests are administered. Schools vary widely in their facilities for the administration of test to groups. This factor was hardly under the control of the investigator.

After administering the test in all the four schools scoring was done with the help of the key. In the case of matching type items and completion type items credit of one was given for each correct response. The formula for correction for guessing was applied in the case of multiple-choice items. The formula is :

$$S=R-\frac{W}{N-1}$$

where

S = corrected score.
R = the number of right responses.
W = the number of wrong responses.
N = the number of possible answers to an item.
(in the present case N = 5).

The Validity and the Reliability of the Final Test

The validity of the test. The empirical relevance of the test has been determined in the present case. School marks of the pupils in their home examination in general science has been taken as the criterion against which empirical relevance of the test is determined. The investigator is well concious that teachers' marks are often unsatisfactory as a criterion because they may be influenced by many other factors than those which should determine scores on achievement test. Such marks are influenced partly by achievement and also by attendance, classroom behaviour, recitations, work habits, general mental ability, socio-economic status, sex, and other aspects of pupils which may or may not be related to achievement test scores. A standardized general science test for the fourth form pupils was not available and hence the teachers marks have been taken as the criterion. The average of the school marks secured by each pupil in physics, chemistry and biology have been calculated. The product moment coefficient of correlation, between the achievement scores in the general science test and the school marks in general science, has been determined.

This correlation coefficient is ·51. Inspite of the unreliability of the teachers marks this validity coefficient is not very low.

It may be pointed out here that due to the inavailability of the I.Q. scores of the pupils the investigator could not take I.Q. scores as the criterion.

The reliability of the final test. The method of 'Rational Equivalence' has been used for determining the reliability of the final test.[12] The advantages of this method have been outlined in the fourth chapter.

The formula for determining the reliability of the test is :

$$\gamma=\left(\frac{n}{n-1}\right)\times\frac{(\text{S.D.}_t)^2-\Sigma xy}{(\text{S.D.})^2}$$

in which

γ = reliability coefficient of the whole test.

n = number of items in the test (in this case 84).

(S.D._t) = the standard deviation of the test scores.

x = the proportion of the group answering a test item correctly.

$y = (1-x)$ = the proportion of the group answering a test item incorrectly.

For determining the value of S.D._t the standard deviation of the scores (correction for guess not being used) has been calculated. The value of S.D._t is 11·69. The chart showing the proportion of pupils passing in each item (x) and proportion of pupils failing in each item (y) has been prepared. For each item 'x' and the corresponding 'y' has been multiplied and their sum is determined which gives Σxy.[13] The value of Σxy is 17·166. On substituting these values in the above formula the reliability coefficient (γ) has been determined. The value of γ is ·88 and the index of reliability ($\gamma_{1\infty}$) is ·94. The index of reliability measures the trustworthiness of test scores by showing how well obtained scores agree with their theoretically true counterparts. The index of reliability gives the maximum correlation which the given test is capable of yielding. This follows from the fact that "the highest possible correlation which can be obtained (except as chance might occasionally lead to higher spurious correlation) between a test and a second measure is with that which truly represents what the test actually measures, that is, the correlation between the test and the true scores of individuals in just such tests."[14] In the present case the index of reliability is ·94 and therefore this is the highest correlation of which this test is capable, since it represents the relationship between obtained test scores and true test scores in the same function.

---

[12] Kuder, G. F. and Richardson, M. W. : *The Theory of Estimation of Test Reliability*, Psychometrika, 1937, 2, 151–160.

[13] See the original thesis.

INTERPRETING TEST DATA

The scores of all the 312 pupils were put in a frequency distribution in the following form.

	→Frequency←							Total
Class-Interval	King Edward's School (Stourbridge)		Technical School (West Bromwich)		Hinckley Grammar School		Holly Lodge Girls Grammar School	Frequency 'f'
	Boys	Boys	Girls	Boys	Girls	Boys	Girls	
70—79	0	0	0	0	0	1	0	1
60—69	2	3	0	0	0	5	0	10
50—59	8	6	0	2	2	12	2	32
40—49	14	14	2	10	10	13	8	71
30—39	10	15	5	15	13	16	20	94
20—29	7	3	6	26	12	3	35	92
10—19	0	0	0	0	0	0	12	12
Total	41	41	13	53	37	50	77	N=312

---

[14] Kelly, T. L.: *The Reliability of Test Scores,* Journal of Educational Research, 1921, 3, 327.

The frequency polygon plotted from the above distribution of 312 scores has been shown in graph No. 3. It is almost a normal curve with a slight positive skewness. The mean (M), the Median (M*d*) and the Standard Deviation (S.D.) of the above distribution are as follows :

Mean (M)=36·3.

Median (M*d*)=35·4.

Standard Deviation (S.D.)=12·06.

It is expected that in the case of an achievement test the average score should be approximately half of the total possible score. In other words it should be fifty per cent. Since the total possible score in this test is 84 and hence mean should have been approximately forty. But due to the low scores of the girls the mean in the present case has been pulled down to 36·3.

The mean, median and standard deviation of the scores for boys and girls have been calculated separately.

Boys mean=40 Boys median=39·1

Girls mean=30·7 Girls median=29·2

Boys Standard Deviation=11·66

Girls Standard Deviation= 9·6

It has been shown in the following way that the difference between the two means is significant.

Sex	Number of Cases (N)	Mean Score (M)	Standard Deviation (S.D.)
Boys	185	40	11·66
Girls	137	30·7	9·6

Standard error of the difference between the two means

$$\text{S.D.}_{(d)}=\sqrt{\frac{(11.66)^2}{185}+\frac{(9{\cdot}6)^2}{137}}$$

$$=1{\cdot}186.$$

The actual difference between the means of boys and girls is 9·3

$$\text{Critical Ratio (C.R.)}=\frac{9{\cdot}3}{1{\cdot}186}=7{\cdot}9.$$

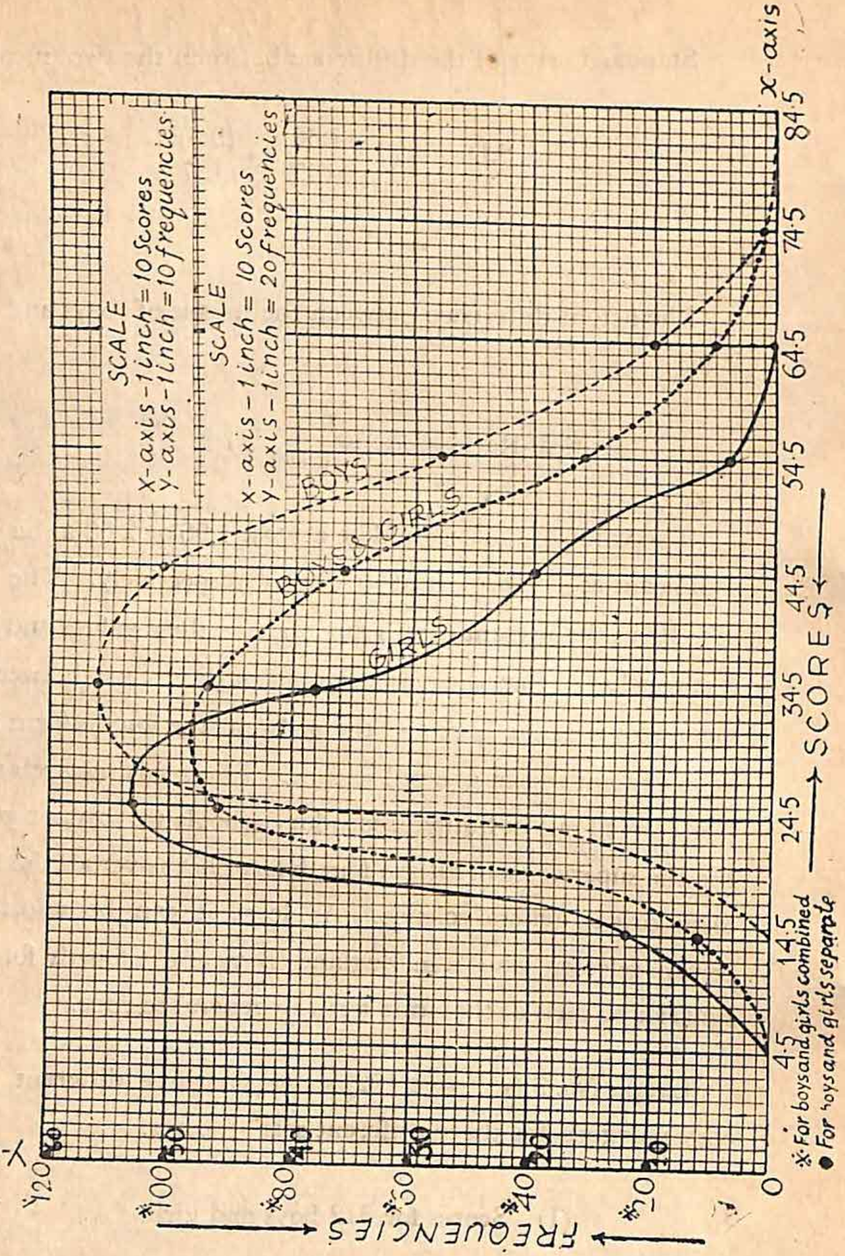
The value of C.R. should be at least 1·96 or 2·58 to be called significant at ·05 level or at ·01 level respectively. The value of C.R. in the present case is much above these values and hence it is concluded that there is a significant difference between the mean scores of the boys and the girls of the fourth form in the achievement test in general science. From this research which was carried out in limited secondary schools we cannot generalise the superiority of all the fourth form boys over all the fourth form girls in general science. At least it can be said that in certain selected schools in England, Boys of the fourth form are superior to girls of the same form in general science.

Graph No. 3 (see page 102) shows the three different frequency polygons for three different sets of scores.

(1) Scores for 312 boys and girls.

(2) Scores for 185 boys.

(3) Scores for 127 girls.



Graph No. 3

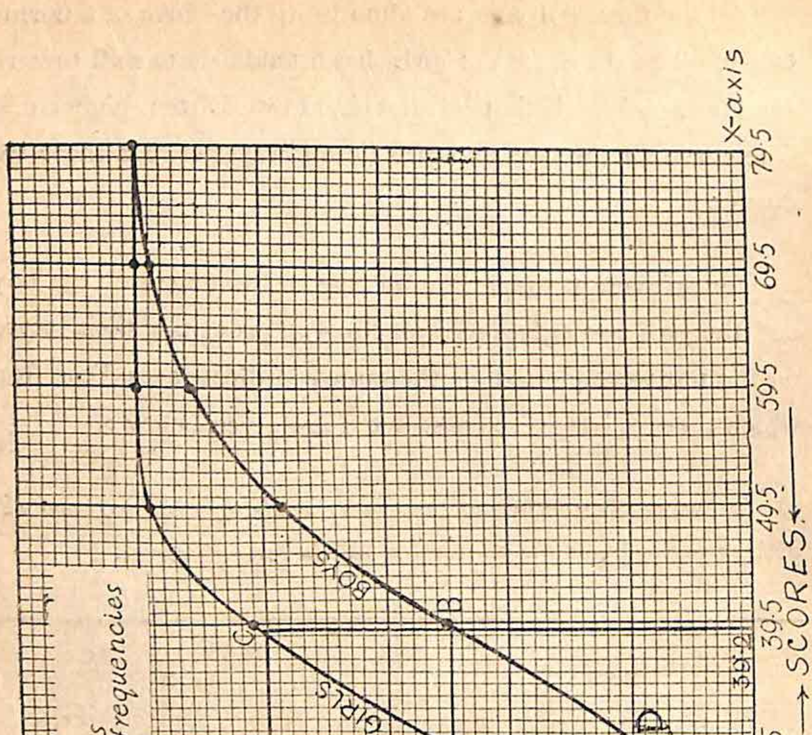
Frequency Polygon of the scores obtained by 185 boys and 127 girls in General Science test. Curve represented by—.—.—.
Frequency Polygon of the scores obtained by 185 boys in the General Science test. Curve represented by..............
Frequency Polygon of the scores obtained by 127 girls in General Science test. Curve represented by————

All the three polygons are almost in the form of a normal curve. The curve for the girls has a tendency to shift towards the lower end and that of the boys towards the higher end. The curve for the boys and girls lie midway between these two curves.

For an overall comparison of the achievement of the boys and the girls two ogives (Graph No. 4) representing their scores on the test were plotted on the same coordinate axes. Data from which these ogives were constructed are given below :

Frequency distributions of the scores, made by 185 boys and 127 girls respectively, on the achievement test in general science.

Boys				Girls			
C. I.	f	Cum. f	Cum% f	C. I.	f	Cum. f	Cum % f
70—79	1	185	100	70—79	0	127	100
60—69	10	184	99·3	60—69	0	127	100
50–59	28	174	93·9	50—59	4	127	100
40—49	51	146	78·8	40—49	20	123	98·4
30—39	56	95	51·3	30—39	38	103	82·4
20–29	39	39	21	20—29	53	65	52
10—19	0	0	0	10–19	12	12	9·6
$N=185$				$N=127$			



Graph No. 4

Ogives representing the scores made by 185 boys and 127 girls in the general science test respectively.

The boys' ogive lies to the right of the girls' over the entire range, showing that the boys score consistently is higher than the girls. It is also clear that differences at the extremes—between the very high-scoring and the very low-scoring boys and girls—are not so great as are differences over the middle range. The boys' median is approximately 39·2, the girls' 29·5 and the difference between these measures is represented by the line AB. By extending the vertical line through B (the boys' median) up to the ogive of the girls' scores, it is clear that approximately 81% of the girls' fall below the boys' median. Hence approximately 19% of the girls exceed the median of the boys in general science test. The vertical line through A (girls' median) cuts the boys' ogive at approximately the 20th percentile. Therefore 20% of the boys fall below the girls' median, and 80% are above this point.

The second problem was to study the comparative achievement of the pupils of the fourth form in general science in the grammar and the technical school. Only one technical school was available and hence the data was inadequate. From this limited data the comparative achievement of all the grammar and technical school pupils cannot be generalised but the superiority or the inferiority of the pupils of a certain technical school in general science to the pupils of certain grammar schools can be predicted. The data was collected from grammar schools and the technical school.

The mean and the standard deviation of the scores of grammar school pupils and of technical school pupils were calculated separately.

Mean score of the grammar school pupils = 37·3

S.D. of the grammar school pupils = 12·3

Mean score of the technical school pupils = 32·1

S.D. of the technical school pupils = 8·5

In order to discover whether these two groups differ significantly in mean achievement in general science the following procedure was adopted.

Schools	Number of cases (N)	Mean Score (M)	Standard Deviation
Grammar School	246	37·3	12·3
Technical School	66	32·1	8·5

Standard error of the difference between the two means

$$= S.D._{(d)} = \sqrt{\frac{(12{\cdot}3)^2}{246} + \frac{(8{\cdot}5)^2}{66}}$$

$$= \sqrt{{\cdot}615 + 1{\cdot}0947}$$

$$= \sqrt{1{\cdot}7097}$$

$$= 1.3075$$

The actual difference between the mean scores of the grammar and technical school pupils is 5·2.

$$\text{Critical Ratio (C.R.)} = \frac{5{\cdot}2}{1{\cdot}3075} = 3{\cdot}9$$

The value of Critical Ratio is more than 1·96 and 2·58 and therefore there is a significant difference between the mean scores of the pupils in certain grammar schools and technical school in achievement test in general science.

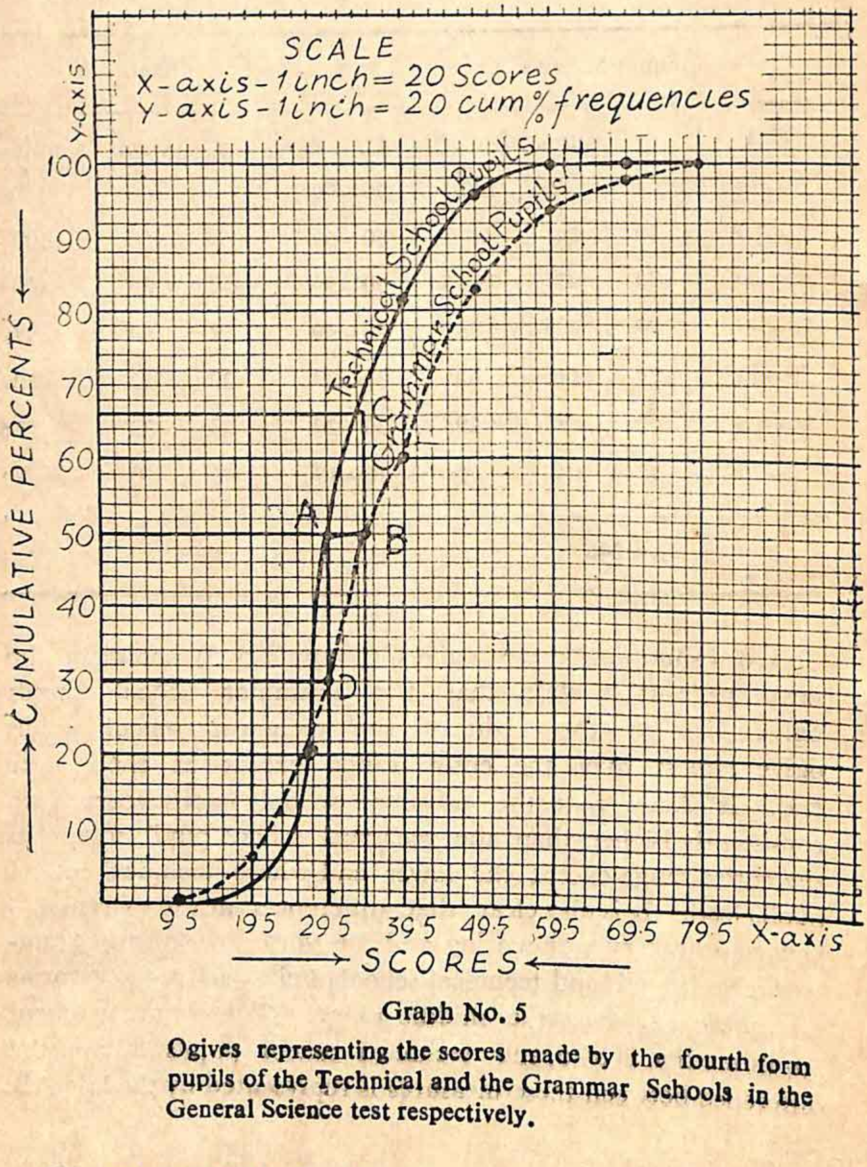
Pupils of the fourth form in certain grammar schools are superior in general science to the pupils of the same form in a particular technical school.

For the graphical representation of the superiority of the grammar school pupils in general science test over the technical school papils two ogives were plotted from the following data :

Grammar Schools				Technical Schools			
C. I.	f	Cum. f	Cum. % f	C. I.	f	Cum. f	Cum. % f
70—79	1	246	100	70—79	0	66	100
60—69	10	245	99·6	60—69	0	66	100
50—59	30	235	95·5	50—59	2	66	100
40—49	59	205	83·7	40—49	12	64	97
30—39	74	146	60	30—39	20	52	80
20—29	60	72	29	20—29	32	32	48·5
10—19	12	12	5	10—19	0	0	0
$N=246$				$N=66$			

On examining these ogives (Graph No. 5) shown in the next page it is clear that the grammar school pupils ogive lies to the right of the technical school pupils ogive almost over the entire range except at the lower end. It shows that the grammar school pupils scores are consistently higher than the technical school pupils scores in the entire range except the lower end where technical school pupils lead. It is also clear that differences at the extremes—between the very high-scoring and the very low-scoring grammar school pupils and technical school pupils—are not so great as are differences over the middle range. The grammar school pupils median is 34·5, the technical school pupils 30 and the difference between these measures is represented by the line AB.

By extending the vertical line through B (the grammar school pupils median) up to the ogive of the technical school pupils scores, it is clear that approximately 66% of the technical school pupils fall below the grammar school pupils median. Hence approximately 34% of the technical school pupils exceed the median of the grammar school pupils in general science test.



Graph No. 5

Ogives representing the scores made by the fourth form pupils of the Technical and the Grammar Schools in the General Science test respectively.

The vertical line through A (the technical school pupils median) cuts the grammar school pupils ogive approximately at the 30th percentile. Therefore 30% of the grammar school pupils fall below the technical school pupils median and 70% are above this point.

Calculation of the average age of the tenth class, first year class, second year calss and fourth form students

On studying the existing pattern of educational system in India it is clear that the expected age of a tenth class pupil in the secondary school is 15+, that of the first year pupil in the intermediate college is 16+, and that of the second year pupil in the intermediate college is 17+. But there are wide variations in the ages of the pupils at each of these stages. On calculating the average ages of the pupils at these different stages respectively it has been proved that the average age resembles the expected age at a particular level. At each stage the average age of those pupils have been calculated who were tested by the investigator in India. The ages of all the pupils who were tested were not available and hence the average age has been calculated from the available ages only. The pupils were encouraged to tell their actual ages. They were convinced that the test would not affect them any way.

Calculation of the average age of 63 pupils of the tenth class in three different High Schools :

Class-interval in terms of months	Frequency f	Deviation in terms of class-interval x'	fx'
216–227	1	4	4
204–215	3	3	9
192–203	7	2	14
180–191	14	1	14
168–179	23	0	0
156–167	11	−1	−11
144–155	4	−2	−8
A.M.=173·5	N=63		$\Sigma fx'=22$

$$M=173{\cdot}5+\frac{22}{63}\times 12=177{\cdot}7 \text{ months}=14 \text{ yrs. } 10 \text{ months.}$$

Calculation of the average age of 108 pupils of the first year class in three different intermediate colleges.

Class-interval in terms of months	Frequency f	Deviation in terms of class-interval x'	fx'
240—251	1	4	4
228—239	1	3	3
216—227	11	2	22
204—215	20	1	20
192—203	29	0	0
180—191	28	—1	—28
168—179	15	—2	—30
156—167	3	—3	—9
	N=108		$\Sigma fx'=-18$

$$\text{A.M.}=197{\cdot}5$$

$$\text{Mean (M)}=197{\cdot}5-\frac{18}{108}\times 12$$

$$=195{\cdot}5 \text{ months}=16 \text{ years } 4 \text{ months.}$$

Calculation of the average age of 160 pupils of the second year class in five different intermediate colleges.

Class-interval in terms of months	Frequency f	Deviation in terms of class-interval x'	fx'
240—251	4	3	12
228—239	14	2	28
216—227	25	1	25
204—215	58	0	0
192—203	41	—1	—41
180—191	17	—2	—34
168—179	1	—3	—3

$$\text{A.M.}=209{\cdot}5$$

$$\text{Mean}=209{\cdot}5-\frac{13}{160}\times 12$$

$$=(209{\cdot}5-{\cdot}97)\text{ months}$$

$$=208{\cdot}03\text{ months}=17\text{ years }4\text{ months.}$$

In England there is a very slight variation in the ages of the pupils at a particular level. The expected age of a pupil in the fourth form of a secondary school is 14+. On calculating the average age of 312 pupils of the secondary schools, who were tested in general science, it comes to 14 years 11 months. The following table illustrates that there is a very slight variation in the ages of the pupils in the fourth form. Most of the pupils lie within the age range of 14 years to 16 years.

Calculation of the average age of 312 pupils of the fourth form in the different grammar and the technical schools.

Class-interval in terms of months	Frequency f	Deviation in terms of class-interval x'	fx'
192—203	6	2	12
180—191	144	1	144
168—179	151	0	0
156—167	10	−1	−10
144—155	1	−2	−2
	N=312		$\Sigma fx'=144$

$$\text{A.M.}=173{\cdot}5$$

$$\text{Mean (M)}=173{\cdot}5+\frac{144}{312}\times 12$$

$$=179\text{ months.}$$

$$=14\text{ years }11\text{ months.}$$

The average age of a tenth class pupil in India is 14 years 10 months and the average age of a pupil in the fourth form of the secondary schools in England is 14 years 11 months. It is therefore concluded that a tenth class pupil of the secondary school in India is comparable to a pupil of the fourth form of the secondary school in England with regard to age. There is a difference of one year in the average ages of the first year pupil in the intermediate college in India and the pupil of the fourth form of the secondary school in England. This difference is of two years in the case of second year pupil in the intermediate college in India and the pupil of the fourth form of the secondary school in England.

THE HINDI VERSION OF THE GENERAL SCIENCE TEST

The English version of the test has been rendered into Hindi to suit the Indian conditions. The scientific terminology having no Hindi Equivalent has been retained in its original form. Though the science subjects are taught through the Hindi medium in the secondary schools and intermediate colleges of Uttar Pradesh yet scientific terminology is taught both in Hindi and English language. In the case of chemistry there are no symbols and formulæ in Hindi language and therefore, chemical formulæ as H_2SO_4, $KMnO_4$, NO etc., were retained in the Hindi version of the test in their original form. Some English terms are easier for the boys to use and pronounce than their Hindi equivalents, *e.g.*—संधारित्र (condenser), क्षुद्र आंत (Ileum), किण्वन (Fermentation) etc. In such cases both the English and the Hindi terms have been given in the Hindi version of the test.

In India the pupils are not very familiar with the New-Type Tests. They find certain difficulties regarding the procedure of taking the test. Sometimes during the course of test programme they need some help regarding the scientific terminology. They need a few minutes of extra time to remove their difficulties. With this end in view the time-limit has been kept as 45 minutes instead of 40 minutes in the Hindi version of the test.

Regarding the justification of the construction of the items it has already been discussed that these items have been constructed keeping in view the common topics that are taught in the fourth form (average age of a pupil 14 years 11 months) of the secondary schools (England) and the tenth class (average age of a pupil 14 years 10 months) of the secondary schools of U. P. (India). The Hindi version of the test claims to measure the same objectives which the English version does.

Administration and scoring of the Hindi version of the test in the Indian Schools (U. P.)

The main purpose of administering the test in the Indian schools was to study the comparative achievement of the pupils, in the Indian and the British secondary schools, in general science.

It has already been pointed out that the pupils of the fourth form in secondary schools in England correspond to the pupils of the tenth class in secondary schools in India in their age. Through the glass of experience the investigator was able to see that the pupils at any particular age level in the English schools are superior, in their achievement in general science, to the pupils at that very particular age level in the Indian schools. The reasons will be discussed later on. However, the Hindi version of the general science test has been applied in the tenth class of three secondary schools in Agra (India). B. R. High School is supposed to be one of the good secondary schools in Agra ; F. C. Inter College is an institution where 70% pupils are from backward classes. On account of this poor stuff the High School results of this institution is always poor. M. L. Girls College is a reputed girls' institution in Agra. 50 boys from B. R. High School, 24 from F. C. Inter College, and 25 girls from M. L. Girls College have been tested in general science. The test was applied in the month of August. All the schools and the colleges in India open in the middle of July each year, after the summer vacations, for the new sessions work. It takes about a fortnight's time to start the routine work in the schools and the colleges due to the heavy rush of admissions. Thus applying the test in the beginning of the session means the assessment of the achievement of the pupils in their previous grades.

Almost the same procedure was followed in presenting the directions as was adopted in the English schools. Pupils in the Indian schools are not very well familiar with the New-Type

Tests and therefore, the procedure for scoring the test was explained more elaborately.

Scoring was done with the help of a key. The same guess correction formula was used as was applied in the case of the English version of the test. Guess correction formula was applied only in the case of the multiple-choice items.

The mean and the standard deviation of the scores have been calculated.

The mean score of the pupils of the fourth form (average age 14 years 11 months) in the grammar and the technical schools in England is 36·3 with a standard deviation of 12·06. The mean score of the pupils of tenth class (average age 14 years 10 months) in U. P. (India) is 21·3 with a standard deviation of 10·2. It has been shown in the following way that the difference between these two means is significant in favour of the pupils of the fourth form in England.

School	Number of cases (N)	Mean Score (M)	Standard Deviation (S.D.)
Pupils of fourth form of the grammar and the technical schools (England)	312	36·3	12·06
Pupils of X class in U. P. (India)	99	21·3	10·2

Standard error of the difference between these two means

$$(\text{S.D.}) = \sqrt{\frac{(12{\cdot}06)^2}{312} + \frac{(10{\cdot}2)^2}{99}}$$

$$= \sqrt{{\cdot}46616 + 1{\cdot}051}$$

$$= 1{\cdot}232.$$

The actual difference between these two means is 15

$$\text{Critical Ratio (C.R.)} = \frac{15}{1{\cdot}232} = 12{\cdot}18$$

The value of 'C.R.' needs to be 1·96 for being called significant at ·05 level and 2·58 for being called significant at ·01 level. In the present case the value of 'C.R.' is much above these values. Therefore, it is concluded that the mean achievement score in general science, of the pupils of the fourth form in the grammar and the technical schools (England) is significantly higher than the mean achievement score of the pupils of tenth class in the secondary schools in U. P. (India).

The Hindi version of the test has been applied to the first year science class of three different intermediate colleges of Agra. 45 pupils of B. R. College (Agra), 43 pupils of M. D. Jain Inter College (Agra) and 20 pupils of Agrawal Inter College (Agra) were tested. The sample was so selected as to contain individuals of widely differing capacities. B. R. College is one of the reputed colleges in the province (Uttar Pradesh). Agrawal Inter College is poor in teaching standard. M. D. Jain Inter College is of medium type. The test has been applied in the last week of August. Scoring was done in the same way as in the case of tenth class pupils. The mean and the standard deviation of the scores has been calculated.

The mean achievement score of the pupils of the fourth form (average age 14 years 11 months) in the grammar and the technical schools in England is 36·3 with a standard deviation of 12·06. The mean score of the pupils of first year science class (average age 16 years 4 months) in U. P. (India) is 30 with a standard deviation of 9·8. It has been shown in the following way that the difference between these two means is significant in favour of the pupils of the fourth form in England.

School	Number of cases (N)	Mean score (M)	Standard Deviation (S.D.)
Pupils of the fourth form of the grammar and the technical schools in England.	312	36·3	12·06
Pupils of first year science class in U. P. (India).	108	30	9·8

Standard error of the difference between these two means :

$$\text{S.D.}_{(\bar{D})}=\sqrt{\frac{(12\cdot06)^2}{312}+\frac{(9\cdot8)^2}{108}}$$

$$=\sqrt{\cdot46+\cdot89}=\sqrt{1\cdot35}$$

$$=1\cdot162$$

The actual difference between these two means is 6·3.

$$\text{Critical Ratio (C.R.)}=\frac{6\cdot3}{1\cdot162}=5\cdot42.$$

This value of 'C.R.' is much higher than 1·96 and 2·58 separately. Therefore, it is concluded that the achievement score in general science of the pupils of the fourth form in the grammar and the technical schools (England) is significantly higher than the mean achievement score of the pupils of first year science class in the secondary schools in U. P. (India).

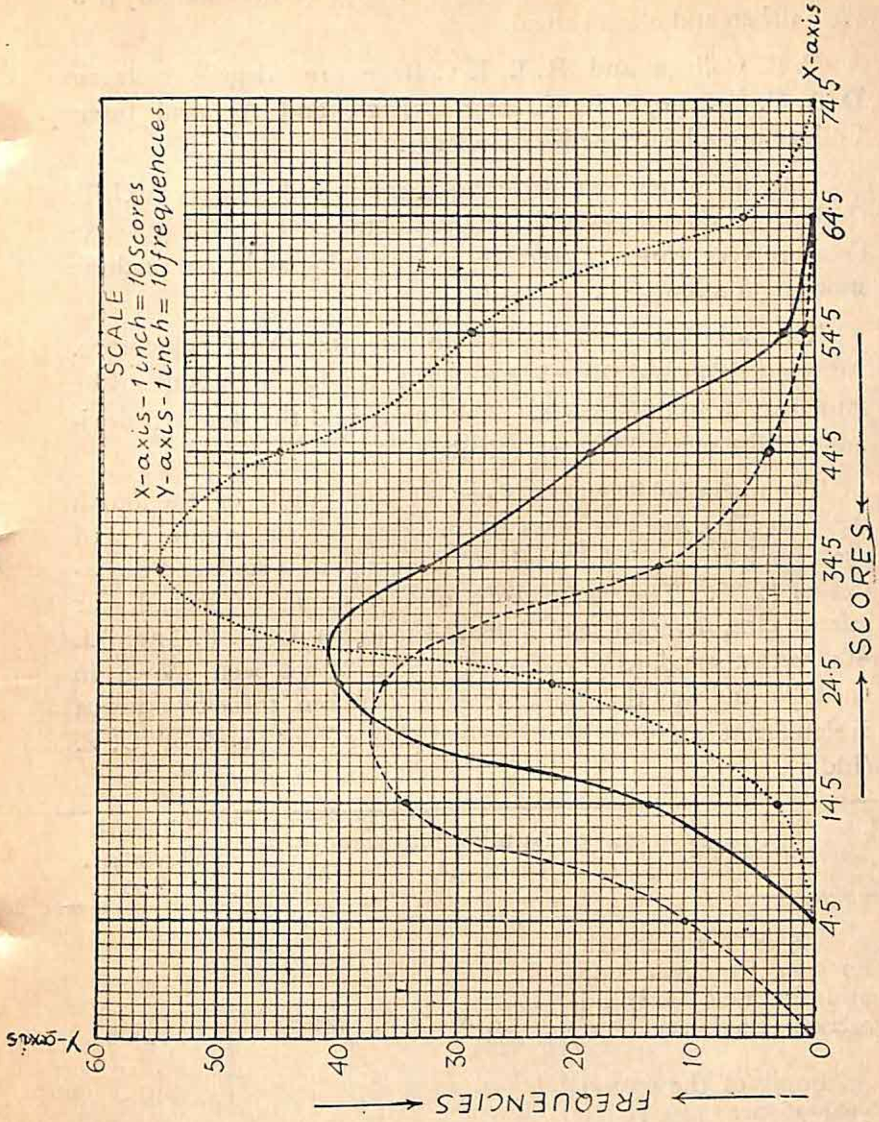
It was thus confirmed that the tenth class pupils and the first year science pupils of U. P. (India) cannot compete the achievement level, in general science, of the fourth form pupils in the technical and the grammar schools in England. The Hindi version of the test was then applied in the second year science class of five different intermediate colleges of U. P. (India). 46 pupils of D. A. V. College (Bulandshahar), 34 pupils of S. R. K. Inter College (Ferozabad), 32 pupils of B. R. College (Agra), 24 pupils of R. E. I. College (Agra), 24 candidates of Agrawal Inter College (Agra) were tested. The following things were kept in view while selecting the sample. Reasons will be discussed later on.

(1) The sample may contain individuals of widely differing capacities.

(2) It may contain pupils both from the rural and the urban areas.

(3) It may contain pupils from the highest classes of the Intermediate Colleges and the pupils from the intermediate sections of the Degree Colleges.

Almost all the pupils in the colleges of Agra are from urban areas. In S. R. K. Inter College (Ferozabad) most of the pupils



Graph No. 6

Frequency Polygon of the Scores made by 160 pupils of the twelfth class in the general science test. Curve represented by- - - - - -

Frequency Polygon of the Scores made by 108 pupils of the eleventh class in the general science test. Curve represented by__________

Frequency Polygon of the Scores made by 99 pupils of the tenth class in the general science test. Curve represented by..........

come from rural areas. D. A. V. College (Bulandshahar) is a semi urban and a semi rural.

B. R. College and R. E. I. College are degree colleges. D. A. V. College, S. R. K. Inter College, and Agrawal Inter College are all Intermediate Colleges.

Only a few per cent girls read in the mixed colleges in U. P. (India). Girls colleges with intermediate science (Biology) sections are very rare. Therefore the sample could not include more than a few girls.

The test has been applied in the month of September. Administration and the scoring of the test has been done in the same way as in other cases. The mean and the standard deviation of the scores has been calculated.

The mean achievement score of the pupils of the fourth form (average age 14 years 11 months) in the grammar and the technical schools in England is 36·3 with a standard deviation of 12·06. The mean score of the pupils of second year science class (average age 17 years 4 months) in U. P. (India) is 40·3 with a standard deviation of 11·04. It has been shown in the following way that the difference between these two means is significant in favour of the second year science pupils of U. P. (India).

School	Number of cases (N)	Mean score (M)	Standard Deviation (S.D.)
Pupils of the fourth form in the grammar and the technical schools (England).	312	36·3	12·06
Pupils of the second year science class (U.P. India).	160	40·3	11·04

Standard error of the difference between these two means :

$$S.D._{(D)}=\sqrt{\frac{(12\cdot06)^2}{312}+\frac{(11\cdot04)^2}{160}}$$

$$= \sqrt{\cdot 4661 + \cdot 76}$$
$$= \sqrt{1 \cdot 2261}$$
$$= 1 \cdot 107$$

The actual difference between these two means is 4.

$$\text{Critical Ratio (C.R.)} = \frac{4}{1 \cdot 107} = 3 \cdot 6.$$

This value of 'C.R.' exceeds both ·05 and ·01 level of significance and hence it is concluded that the achievement score, in general science, of the pupils of second year science (Biology) class in U. P. (India) is significantly higher than that of the pupils in the fourth form of the grammar and the technical schools in England.

The mean achievement score of the pupils (excluding the girls) of the fourth form in the grammar and technical schools in England is 39·1 with a standard deviation of 11·66. The sample for the second year science pupils of U.P. (India) included almost all the boys with the exception of a few girls. The mean achievement score is 40·3 with a standard deviation of 11·04. It has been shown in the following way that the difference between these two mean achievement scores is not significant.

School	Number of cases (N)	Mean score (M)	Standard Deviation (S.D.)
Boys of the fourth form in the grammar and the technical schools in England.	185	39·1	11·66
Boys of the second year science in U. P. (India).	160	40·3	11·04

Standard error of the difference between these two means :

$$\text{S.D.}_{(D)} = \sqrt{\frac{(11 \cdot 66)^2}{185} + \frac{(11 \cdot 04)^2}{160}}$$

$$= \sqrt{\cdot 7359 + \cdot 76}$$

$$= 1 \cdot 223.$$

The actual difference between these two means is 1·2.

$$\text{Critical Ratio (C.R.)} = \frac{1 \cdot 2}{1 \cdot 223} = \cdot 97.$$

This value of C.R. is much below ·05 level of significance and ·01 level of significance and hence it is concluded that the mean achievement score, in general science, of the boys of the fourth form in the grammar and the technical schools in England is same as to that of the second year science boys in U.P. (India).

For the graphical representation of the comparative general science achievement of the pupils in the various grades of secondary schools (U. P.) and the fourth form of secondary schools (England) six ogives were plotted from the following data :

X Sc. (U. P.)				XI Sc. (U.P.)				XII Sc. (U.P.)			
C.I.	*f*	Cum. *f*	Cum. %	C I.	*f*	Cum. *f*	Cum. %	C.I.	*f*	Cum. *f*	Cum. %
50–59	1	99	100	50–59	2	108	100	60-69	6	160	100
40–49	4	98	99	40–49	19	06	98 1	50–59	29	154	96·3
30–39	13	94	95	30–39	33	87	80	40-49	45	125	78
20–29	36	81	82	20–29	40	54	50	30-39	55	80	50
10–19	34	45	45.5	10–19	14	14	13	20-29	22	25	15·6
0–9	11	11	11	0–9	0	0	0	10-19	3	3	2
N=99				N=108				N=160			

Grammar School pupils (England) Fourth Form				Technical school pupils (England) Fourth Form				Pupils (excluding the girls) of the grammar and the technical schools Fourth Form			
C.I.	*f*	Cum.*f*	Cum. %	C.I.	*f*	Cum.*f*	Cum. %	C.I.	*f*	Cum.*f*	Cum. %
70–79	1	246	100	70–79	0	66	100	70–79	1	185	100
60–69	10	245	99·6	60–69	0	66	100	60–69	10	184	99·3
50–59	30	235	95·5	50–59	2	66	100	50–59	28	174	93·9
40–49	59	205	83·7	40–49	12	64	97	40–49	51	146	78·8
30–39	74	146	60	30–39	20	52	80	30–39	56	95	51·3
20–29	60	72	29	20–29	32	32	48 5	20–29	39	39	21
10–19	12	12	5	10–19	0	0	0	10–19	0	0	0
N=246				N=66							

Graph No. 7 given on the next page shows the six ogives representing scores made by the following groups of pupils :

(1) X Sc. pupils (U. P.) ; (2) XI Sc. pupils (U. P.) ; (3) XII Sc. pupils (U. P.) ; (4) Grammar school pupils (fourth form) ; (5) Technical school pupils (fourth form) ; (6) Pupils (excluding the girls) of the fourth form in the grammar and the technical schools.

On examining these ogives the following conclusions are drawn :

(1) The ogives of XII Sc. pupils (U. P.) and the pupils (excluding the girls) of the fourth form in the grammar and the technical schools (England) almost overlap each other showing that their achievement in general science is equal.

(2) The ogives of XI Sc. pupils (U. P.) and pupils of the technical schools (England) overlap showing that their achievement in general science is equal.

(3) The ogives of X Sc. pupils (U. P.) and XI Sc. pupils (U. P.) lies to the left of the ogive of the fourth form pupils in the grammar and the technical schools (England) showing that

$$\text{Critical Ratio (C.R.)} = \frac{D}{S.D._{(D)}} = \frac{\cdot 13}{2\cdot 04} = \cdot 068.$$

This value of C.R. (·068) is much below ·05 and ·01 level of significance and hence it is concluded that there is no significant difference between the general science achievement of the XII Sc. pupils reading in the degree colleges and that of the pupils reading in the intermediate colleges.

The sample for the XII Sc. pupils consists of 56 pupils from rural areas and 104 pupils from urban areas. This sample is evidently inadequate to draw any definite conclusions regarding the difference between the mean performance of the two groups. To have at least a rough idea of the comparative science achievement of the XII Sc. pupils from the rural and urban areas the following procedure was followed.

The mean and the standard deviation of the scores of rural area pupils and urban area pupils were calculated separately.

The mean and the standard deviation of the scores of rural and urban area pupils are shown in the following table :

Pupils	Number 'N'	Mean 'M'	Standard Deviation (S.D.)
Pupils from rural areas	56	39·9	11·8
Pupils from urban areas	104	40·56	10·6

$$\text{Actual difference between means} = D = 40\cdot 56 - 39\cdot 9$$
$$= \cdot 66$$

Standard error of the difference between the mean scores

$$S.D._{(D)} = \sqrt{\frac{(11\cdot 8)^2}{56} + \frac{(10\cdot 6)^2}{104}}$$
$$= \sqrt{4\cdot 56}$$
$$= 2\cdot 14$$

$$\text{Critical Ratio (C.R.)} = \frac{D}{S.D._{(D)}} = \frac{\cdot 66}{2\cdot 14} = \cdot 308.$$

This value of C.R. (·308) is much below ·05 and ·01 level of significance and hence it is concluded that there is no significant difference between the general science achievement of the XII Sc. pupils from rural areas and from urban areas.

The reliability of the Hindi version of the test

The second year science students in India almost match the fourth form pupils in the secondary schools in England in general science achievement. It was, therefore, thought proper to calculate the reliability coefficient of the Hindi version of the general science test with the data for the second year science class only.

The method of 'Rational Equivalence' has been used for determining the reliability of the Hindi version of the test.

The formula for determining the reliability of the test is :

$$\gamma=\left(\frac{n}{n-1}\right)\times\frac{\text{S.D.}_t^2-\Sigma pq}{\text{S.D.}_t^2}$$

in which

γ = reliability coefficient of the whole test.

n = number of items in the test (in this case 84).

S.D._t = the standard deviation of the test scores.

p = the proportion of the group answering a test item correctly.

$q=(1-p)$ = the proportion of the group answering a test item incorrectly.

For determining the value of S.D._t, the standard deviation of the scores (correction for guess not being used), has been calculated.

The value of S.D._t is 10·87. The chart has been prepared to show the proportion of pupils passing in each item (x) and proportion of pupils failing in each item (y). For each item 'x' and the corresponding 'y' have been multiplied and their sum is determined which gives Σxy. The value of Σxy is 16·7061.

On substituting these values in the formula the reliability coefficient has been calculated.

$$\text{reliability coeff. } (\gamma)=\frac{n}{n-1}\times\frac{\text{S.D.}^2_t-\Sigma pq}{\text{S.D.}_t^2}$$

$$=\frac{84}{83}\left[1-\frac{16\cdot7061}{(10\cdot87)^2}\right]$$

$$=1\cdot01[1-\cdot14]$$

$$=1\cdot01\times\cdot86$$

$$=\cdot8686$$

Index of reliability $=\gamma_{1\infty}$

$$=\sqrt{\cdot8686}$$

$$=\cdot93.$$

This index of reliability of the test indicates that the highest correlation of which the Hindi version of the test is capable is ·94 and this is sufficiently high for our purpose.

The validity of the Hindi version of the test

The empirical relevance of the Hindi version of the test has been estimated in the following ways.

The determination of the empirical relevance of a test may include the differences in average test scores obtained by any two or more groups known to be widely separated according to ages or grades. The Hindi version of the test has been applied in the tenth class, first year science class and second year science class pupils. The following table shows their respective average ages, test scores and standard deviations.

	Average age	Mean achievement score in general science test	Standard deviation 'σ' of test scores	No. of Pupils
Tenth class pupils	14 years 10 months	21·3	10·2	99
First year science pupils	16 years 4 months	30	9·8	108
Second year science pupils	17 years 4 months	40·3	11·04	160

The following table shows that the mean achievement score of the second year science pupils is significantly higher than that of the tenth class pupils and the first year science pupils. The mean achievement score of the XI year science pupils is significantly higher than that of X class pupils.

	Difference in mean achievement score.		Standard error of the difference between means (σ_D)		Critical Ratio		Interpretation
XII vs. X	·84		1·345		·63		Difference is significant in favour of XII year pupils.
XII vs. XI	10·3		1·285		8·01		Difference is significant in favour of XII year pupils.
XI vs. X	8·7		1.714		5·07		Difference is significant in favour of XI year pupils.

Thus it is clear that the mean achievement score of the pupils at the higher grades is significantly greater than the mean achievement scores of the pupils at the lower grades. It is evidently a good index of empirical validity of the test.

Another method was also followed to determine the empirical relevance of the test for the second year science class. The school marks of the pupils, of the second year science, in physics, chemistry and biology were taken. Since the teachers' marks are often unsatisfactory as a criterion because they may be influenced by many other factors than those which should determine scores on achievement test, therefore average marks of the pupils in three home examinations [First year science class (Annual examination) ; Second year science class (first and second terminal) examination] were calculated to be taken as a criterion. The product moment coefficient of correlation between the achievement scores in the general science test and the average school marks in general science has been determined.

The school marks of only 110 pupils were available. The coefficient of correlation between the achievement score in general science and the corresponding average school marks in general science is ·52.

This shows that there is a high correlation between the average school marks and the achievement scores in general science secured by 110 pupils of the second year science class. It means that the test measures the science ability of the pupils efficiently.

INTERPRETATIONS

This study shows that the mean general science achievement score of the pupils of the fourth form (average age 14 years 11 months) in the chosen grammar and the technical schools in England is equivalent to the mean general science achievement score of the pupils of the second year science (average age 17 years 4 months) in the chosen Intermediate Colleges of U. P. (India). Thus there is a difference of 2½ years in the general science achievement level of the pupils in the chosen secondary schools of England and India (U. P.). The following may be the reasons for this difference.

The first and the foremost reason for this difference is that the curriculum of science in the secondary schools of India (U. P.) is not rich. Most of the things which the pupil comes across in his earlier life are taught at the later stage. In almost all Indian homes (in urban area) electricity is used but up to high school stage pupils are not taught the units of current, potential difference, etc. They cannot take the electric meter reading correctly. In contrast to this I have seen British children, reading in the secondary schools, who can construct a wireless set. They get such an advanced knowledge of science not only in their schools but their homes are very inspiring. Science is a study of the problems that are found wherever children live. It is a study of natural environment—not merely pieces of chemistry, physics and biology and astronomy and geology. If science is to be learned effectively it must be experienced. It must be learned and not learned about.

Good science teaching must be based on observation and experiment. But performing experiment and learning to make close observations require special facilities and well equipped laboratories and these are lacking in India, especially at the elementary and secondary levels. The laboratory is a place where pupils get first hand experience with materials that may or may not be new to them, but which are used with one end in view, the answering of a question which can only be answered by contact with natural objects. Using the laboratory develops better understanding and interpretations of the principles of science. In the laboratory the pupil himself proves, verifies and establishes the principles which have long since been well established and generally accepted. These principles which the pupil verifies for himself are learnt by him permanently. It is not an exaggeration to say that practical work has an enduring effect on the understanding of theory. During the course of practical work the child faces the real problems. Laboratory problems must arise as a significant need from the class-room discussion. Laboratory problems and class-room problems should be the same, differing only in method of approach—the situations being arrived at through the same means. A real understanding of science depends upon the proper correlation between the laboratory period and the class-room period.

It is shocking that at present there is no scope for laboratory work in the secondary schools of India. Some experiments have been prescribed for high school course but neither the teachers nor the pupils take real interest in experimentation. The reason is that no external examination is conducted up to tenth class to assess the practical work done by the pupils during the session. It is suggested that some means of assessment of the practical work of the pupils must be used so that they may take a genuine interest in it. Lack of the proper equipment of the science laboratories is also a handicap for Indian school children. Now the National Council of Educational Research and Training is paying a right attention toward this direction.

All science experiments are not necessarily to be done in the laboratory. There are experiments which pupils can perform at their homes. In England teachers encourage the pupils to do experiments at home, and receive further guidance in the schools. In India neither the teacher encourages the child to perform any experiment at home nor do they get any opportunity in the homes for doing so because of lack of the proper home environment. It is, therefore, English pupils of a particular age are far superior to Indian pupils of the same age in general science.

The second reason is that in each grade of Indian schools there is a wide variation in the ages, abilities and aptitudes of the pupils. The class is not a homogeneous group. The same instructions, the same syllabus, and the same time-table is followed for the bright pupils and the dull ones. After all we cannot fit square pegs into round holes. There are many advantages of segregating the pupils in the secondary schools according to their age, ability and aptitude.

Homogeneity in grouping simplifies the teacher's problems and promotes instructional efficiency. It simplifies organization and internal administration; "for there is one usually one standard 'diet' and one form of 'service'. The chief remedy (there are others) for the 'indigestibility' of the food is a reduction in the quantity rather than a change in the menu."[15]

If the group is a heterogeneous one, the ablest pupils are bound either to strike the distinctive note or else to suffer from insufficient stimulus and direction of their powers. In the latter case the interests of individuals and of the community would both suffer; and in the former case serious injustice would be done to many whose capacities are least likely to be fostered in a 'mileu' where intellectual achievement is the approved goal of effort and inability to reach it is failure. The differences that are to be considered in segregating the pupils to different groups or different schools (as is the practice in England) are not only the differences in the 'type of mind' but the differences in general intelligence too. "At the critical age of eleven the dullest children attending primary schools in England have been found to have a mental age of eight years, the brightest a mental age of fourteen or more".[16] If there is any truth in this finding then we cannot think of imposing upon the duller children a curriculum that would constantly baffle their modest abilities and rob them of the pleasure they might and ought to feel in tasks within their powers, or to hold back the brighter ones from the high intellectual enterprises nature has fitted them to undertake.

---

[15] Earle, F. M.: *Reconstruction in the Secondary School,* University of London Press Ltd., Malham House, Bickley, Kent, 1945.

In England, according to 1944 act, every child receives the education appropriate to his age, ability and aptitude. At the age of eleven plus children are directed from the primary schools to secondary schooling best suited to their needs and capacities. For this purpose Intelligence test, English test, Arithmetic test and Cumulative record cards are used. Borderline cases are interviewed. The different types of secondary schools provide the curricula that suits the needs and capacities of different pupils attending them. Thus the standard of work in grammar schools is maintained at a high level ; for on the maintenance of standards in the schools depends in large measure the quality of the studies pursued in universities and other places of higher education. The Secondary Modern Schools provide a curricula closely related to the immediate interests and environment of its pupils and adopt a method of approach which is practical and concrete. In the Technical Schools the curriculum for pupils between the ages of 11+ and 13+ is broadly of the same character as the curriculum in other types of secondary schools of equal status. For pupils of 13+ and onwards the curriculum is designed to provide a liberal education with science and its applications as the core and inspiration.

General science is taught in every type of secondary schools in England but the method of treatment and content of subjects differ according to the type of mind and general intelligence of the pupils attending them.

[14] Burt, C. : *The Education of the Young Adolescent*, Brit. Journal of Educational Psychology, Nov., 1943.

Many studies in England have been carried out which convince us that special interest and abilities do in fact often reveal themselves clearly by the age of 10+ or 11+ and hence classification of pupils is done at 11+. But there is another age point, namely 13+, in all schools at which transfer is made, if desirable, from one type of school to another with the object both of eliminating any misfits which may occur, and with the purpose of encouraging any special aptitudes which may develop in children between the ages of 11+ and 13+. Children are taught common type of courses between the ages 11+ and 13+ so that there may not be any difficulty in transferring children from one type of school to another type. It will be an exaggeration to say that all the secondary schools follow this method strictly. At present, on an average, one per cent pupils are transferred from one type of school to another type.

In England though the pupils are segregated to different schools according to their age, ability and aptitude yet there is no scientific method of selecting the pupils for the general science course in these schools. It mostly depends upon the teachers estimate and cumulative record cards. The mean general science score of 312 pupils in the secondary schools of England is 36·3 (43%) with a standard deviation of 12·06. It shows that there are wide variations between pupils achievement in general science. Some sort of predictive tool is necessary for the proper selection of the pupils for the general science course.

The third reason for this difference between the achievement of the Indian and English pupils may be the lack of proper scientific terminology in Hindi. This is a transition period in India. Most of the scientific terms have been translated into Hindi but the difficulty is that many of the teachers who have graduated 10 years back follow the English scientific terminology. They find it difficult to teach through the medium of Hindi. But in the Higher Secondary stage Hindi medium is compulsory in U .P. Thus in many cases pupils are left half-way. The following illustration will make the point clear.

Only 13% pupils in the tenth class could answer the following item correctly :

The teeth whose function is tearing flesh are known as (...........).

Out of these 13% pupils 6% gave the answer as 'कैनाइन' (canine) but only 2% spelled 'कैनाइन' (canine) correctly. 3% answered it as 'कील दाँत' and 4% as 'छेदन दाँत'.

It shows that there is no uniformity in some of the scientific terminology in Hindi. It is suggested that new books should be introduced in which technical words to be used should conform to the list published by the Government of India. An abridged list of these technical terms should be included in these books for ready reference.

The second section of 'Part I' of the test has been used to test pupils ability of understanding of units. On an average 15% pupils of tenth class (U.P.), 24% pupils of XI science class (U. P.), 43% pupils of XII science class (U. P.) and 42% pupils of the fourth form of the grammar and the technical schools (England) could answer this section correctly. It shows that very little emphasis is given to the units up to the High School stage in the schools of U. P. (India). It is not too much to familiarize the tenth class pupil with the units of a power of a lense, electric current, work, potential difference, force, pressure, power, etc. A child can express himself correctly when he has a clear conception of the units.

A fourth reason may be the lack of the training of the teachers of science in India. It is one thing to be a good scientist and another to be a good science teacher. On account of the new advances, the curriculum of schools is constantly, undergoing change. If the teacher is to remain effective in such a changing scene his competence must increase correspondingly. The academic and professional training of teachers has been so far of a formal nature, consequently, the teaching by them tends to be equally formal. They have lacked much of the competence essential for the encouragement of and provision for genuine experiences. An Indian science teacher is always busy in completing his course which is prescribed by the Board of Education or the University. He does not care whether the children are benefitted by it or not. In England there is a freedom to the teacher for determining the curriculum and methods of teaching. Neither the Minister nor the Local Education Authority exercises authority over the curriculum of any school beyond that of agreeing the general educational character of the school and its place in local educational system. The governers in consultation with the head of the school, are normally responsible for the general direction of the curriculum of the school but in practice the curriculum is settled by the head in co-operation with the assistant staff. Naturally the pupils of that country will achieve more who are taught according to their needs and not according to the needs of the authorities.

Time table plays an important role in teaching. In India time table is not framed according to the need of science. There is never a provision of preparatory period for making arrangements and preparations for demonstration and experimentation.

In the Indian schools we have overweighted some of the school subjects and consequently general science has been underweighted. Mathematics has the most prominent place in the school curriculum. It is a compulsory subject up to the tenth class but general science is compulsory only up to the eighth class. It may be the reason as to why pupils in India are good in mathematics but not so in general science. In England much importance has been given to general science as a school subject.

There are 11 items in the test which have been used to obtain the evidence of pupils ability to apply scientific principles to situations new to students. The following table shows the percentage of the pupils who are able to respond correctly to those items.

Item No.	XII science class (U.P.) Percentage of pupils answering correctly	Fourth form pupils (England) Percentage of pupils answering correctly
2	28	48
3	77	85
8	44	46
10	31	23
18	18	29
19	52	47
21	47	57
22	27	57
24	27	13
4	51	41
5	17	14

It is curious to note that on an average 38% pupils in the schools of U. P. and 42% pupils in the schools of England could answer these items correctly. It seems that the application side of the pupils is weak. The fault may be of teaching. I can easily say that in the Indian schools science is becoming a formal routine of experiments whose import in life fails to be clearly brought out ; indeed there is much to be said for the method of teaching which starts with applications and proceeds to principles instead of following the reverse order.

(*Pilot-Test*)

GENERAL SCIENCE TEST

Please print

Surname.................................
(in capital letters)
Christian Names...........................
(in capital letters)
Name of School.............................
Your age.........years...........months. Your sex...........
(Boy or Girl)

Number of years you have studied general science or science so far..

General Directions

(1) You may use the other side of this sheet for your rough work.

(2) Underline any words you do not understand.

(3) Do not turn this page until the examiner tells you to do so.

(4) This test consists of four parts and requires 80 minutes of working time.

(5) The directions for each part are printed at the beginning of the part. Read them carefully and proceed at once to answer the questions.

(6) Do not spend too much time on any one item. Answer the easier questions first, then return to the harder ones if you have time.

(7) There is no time limit for each part.

(8) You are expected to attempt as many questions as you can from each part.

(9) Guessing will reduce your score, so do not guess.

(10) Use pencil for answering. If you want to change your answers, do so by rubbing out and rewriting them.

PART 1 : DIRECTIONS

In most of the following sections there are two columns. For each statement or question in the first column there is an answer in the second column. Find the answer and place its letter in the brackets in front of the item in the first column. The first item in Section I is done for you as an example. Answer as many questions as you can in all eight sections in the same way.

Section I

What is the chemical formula for one molecule of :

(I) (1) nitric oxide
( ) (2) ammonium chloride
( ) (3) nitrous oxide
( ) (4) nitric acid
( ) (5) calcium hydroxide
( ) (6) potassium permanganate
( ) (7) sulphuric acid
( ) (8) hydrogen sulphide (or sulphurated hydrogen)
( ) (9) ammonium sulphate
( ) (10) sodium chloride

(A) H_2SO_4
(B) $KMnO_4$
(C) NACl
(D) $CACO_3$
(E) $(NH_4)_2SO_4$
(F) HNO_3
(G) NH_4OH
(H) NH_4Cl
(I) H_2S
(J) N_2O
(K) $Ca(OH)_2$
(L) NO

Section II

What units are the following measured in :

( ) (1) density
( ) (2) power of a lens
( ) (3) latent heat
( ) (4) work
( ) (5) quantity of heat
( ) (6) potential difference
( ) (7) force
( ) (8) power
( ) (9) pressure
( ) (10) electric current

(A) ft. lb.
(B) cm. per sec. per sec.
(C) ampere
(D) volt
(E) dyne
(F) foot pound per sec.
(G) grams per c. c.
(H) calories
(I) dioptre
(J) calories per gram
(K) pound weight
(L) pound weight per square inch

Section III

On the left hand side there are ten processes. On the right hand side there are ten descriptions not in the same order. In the brackets provided write the letter which indicates the statement corresponding to each process in turn.

( ) (1) conduction
( ) (2) electrolysis
( ) (3) filtration
( ) (4) convection
( ) (5) refraction
( ) (6) distillation
( ) (7) radiation
( ) (8) evaporation
( ) (9) plasmolysis
( ) (10) osmosis

(A) The method used for separating an insoluble substance from a liquid.

(B) The bending of a ray of light which travels obliquely from one transparent medium to another.

(C) The way in which heat is carried from one point to another in solids.

(D) The method used for separating a dissolved substance from a liquid.

(E) The main way in which heat is carried from one point to another in a liquid.

(F) The collapse and contraction of the protoplasm of a cell due to the passing of water out of the cell.

(G) The process by which liquids and substances in solution tend to pass through permeable and semi-permeable membrane.

(H) The process of splitting up a substance chemically by means of an electric current.

(I) A process used for obtaining pure water from a solution.

(J) The way in which heat reaches us from any source without appreciably warming the intervening medium.

Section IV

What is the chemical name of each of the common names given in the first column ?

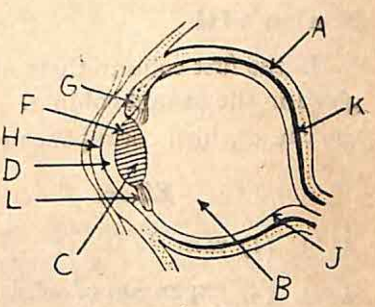
( ) (1) chalk
( ) (2) slaked lime
( ) (3) blue vitriol
( ) (4) calamine
( ) (5) washing soda
( ) (6) nitre
( ) (7) lime stone
( ) (8) gypsum
( ) (9) common salt
( ) (10) caustic potash

(A) copper sulphate
(B) sodium chloride
(C) zinc sulphate
(D) calcium carbonate
(E) potassium hydroxide
(F) calcium sulphate
(G) sodium carbonate
(H) sodium hydroxide
(I) calcium hydroxide
(J) sodium nitrate
(K) zinc carbonate
(L) potassium nitrate

Section V

The following diagram shows the section of an eye. Which pointers in the diagram indicate the following parts :

( ) (1) sclerotic
( ) (2) cornea
( ) (3) iris
( ) (4) pupil
( ) (5) aqueous humour
( ) (6) vitrous humour
( ) (7) optic nerves
( ) (8) ciliary muscles
( ) (9) retina
( ) (10) lense



Section VI

Place the letter of the colour of each of the following sulphides in the bracket in front of it.

( ) (1) copper sulphide
( ) (2) stannous sulphide
( ) (3) antimony sulphide
( ) (4) arsenic sulphide

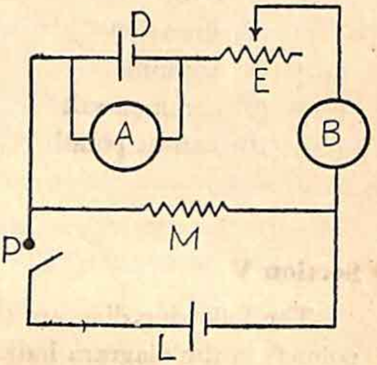
(A) yellow
(B) orange
(C) black
(D) white

( ) (5) zinc sulphide	(E) brown
( ) (6) manganese sulphide	(F) buff
	(G) red
	(H) green

Section VII

The following diagram represents an electric circuit. Which pointer indicates the following :

( ) (1) voltmeter
( ) (2) ammeter
( ) (3) variable resistance
( ) (4) bettery
( ) (5) fixed resistance
( ) (6) key
( ) (7) condenser



Section VIII

In the first column there are some effects. Their causes are given in the second column. In the brackets write the letter of the cause which corresponds to its effect.

Effect	**Cause**
( ) (1) rain	(A) Total reflection of light.
( ) (2) expansion of solids	(B) The fall of air temperature below the saturation point.
( ) (3) mirage	(C) Gravitational force.
( ) (4) weight	(D) Low atmospheric pressure.
( ) (5) dew	(E) Condensation of water vapour on small particles in the atmosphere.
( ) (6) fog	(F) Rise in temperature.

PART 2 : DIRECTIONS

Each of the following items is followed by five answers. Select the best possible answer in each case and put its letter in the bracket on the left of the item. Item (1) is solved for you as an example.

(1) The apparatus used for measuring the atmospheric pressure is called :

(b) (a) thermometer (b) barometer
(c) pyrometer (d) lactometer
(e) hydrometer

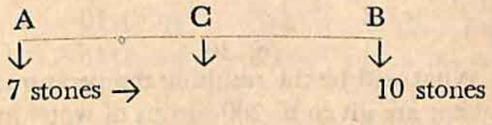
(2) The weight of one cu. ft. of water is 62·5 lbs. and one cu. ft. of cork is 15 lbs. The specific gravity (relative density) of the cork is :

( ) (a) 4·17 (b) ·24
(c) ·417 (d) 2·4
(e) 15

(3) The weight of a piece of iron in air is 10 lbs. and its apparent weight in water is 8·7 lbs. Its specific gravity is :

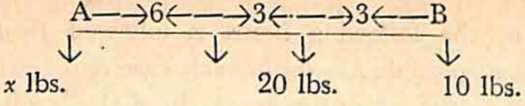
( ) (a) ·13 (b) 1·87
(c) 7·7 (d) 8·7
(e) 10

(4) A boy weighing 10 stones sits on one end (B) of a light 10 ft. plank. Another boy weighing 7 stones sits at the other end (A). The plank is supported by a single trestle (C). At what distance from (A) must it be placed so that the plank balances ?



( ) (a) 3 feet (b) ·7 feet
(c) $5\frac{15}{17}$ feet (d) $4\frac{2}{17}$ feet
(e) 1 foot

(5) If the rod AB (figure below) balances, then x is equal to:



( ) (a) 20 lbs. (b) 30 lbs.
(c) 10 lbs. (d) 15 lbs.
(e) 5 lbs.

(6) Which of the following are physical (not chemical) changes :

( ) (a) rusting of iron (b) change in magnesium wire when heated
(c) change in platinum wire when heated (d) the dissolving of copper sulphate in water
(e) that represented by the equation
acid + base = salt

(7) The mouth of a vacuum flask is closed with a cork because :

( ) (a) cork is cheap (b) cork is durable
(c) cork is washable (d) cork is a bad conductor of heat
(e) cork is lighter in weight than other substances

(8) In cold climates woollen clothes are used because :

( ) (a) they are durable (b) they give heat to the body
(c) they diminish the loss of heat from the body (d) they are heavier than cotton clothes
(e) they are good looking

(9) How many British Thermal units are required to raise 10 pounds of water through 20 degrees Fahrenheit ?

( ) (a) 200 (b) 20
(c) 2 (d) 10
(e) 30

(10) What will be the resulting temperature if 500 calories of heat are given to 200 grams of water at 10^0C ?

( ) (a) 17^0C (b) $12{\cdot}5^0$C
(c) $10{\cdot}4^0$C (d) 13^0C
(e) $7{\cdot}5^0$C

(11) How much heat is given out by a copper ball of mass 200 gm. in cooling from 300°C to 0°C, if the specific heat of copper is 0·1 cal. per gram per degree C ?

( ) (a) 50 calories (b) 100 calories
(c) 500 calories (d) 60,000 calories
(e) 6,000 calories

(12) The latent heat of ice is 80 calories per gram. How much heat is required to change 10 gm. of ice at 0°C to water at 30°C ?

( ) (a) 1,100 calories (b) 240,000 calories
(c) 500 calories (d) 380 calories
(e) 770 calories

(13) Sometimes we treat water with chlorine gas because :

( ) (a) it makes water colourless (b) it helps the floating particles on the water to settle down
(c) it removes the bad smell of water (d) it kills harmful germs
(e) it adds taste to the water

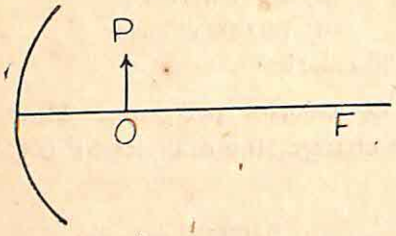
(14) The increase in length per unit of length of a metal rod for each degree rise in temperature is known as :

( ) (a) the specific heat of the metal (b) the elasticity of the metal
(c) the coefficient of linear expansion of the metal (d) the surface tension of the metal
(e) the coefficient of volume expansion of the metal

(15) Two plane mirrors are inclined at an angle of 60°. The number of images of an object placed between them will be :

( ) (a) 5 (b) 6
(c) 7 (d) 3
(e) 2

(16) The diagram shows an object OP in front of a concave mirror between the principal focus (F) and the mirror. The image formed will be :



( )
(a) the same size as object, inverted, real
(b) magnified, erect, real
(c) magnified, inverted, real
(d) magnified, erect, virtual
(e) smaller than object, inverted, real

(17) The focal length of a concave mirror is 9 cm. An object is placed at a distance of 27 cm. from the mirror. The distance of the image from the mirror will be :

( )
(a) 9 cm.
(b) 13·5 cm.
(c) 27 cm.
(d) 6·75 cm.
(e) 18 cm.

(18) It is **not** true that the image formed by a convex mirror is always :

( )
(a) virtual
(b) diminished
(c) erect
(d) the same distance behind the mirror as the object is in front
(e) not at the same distance behind the mirror as the object is in front

(19) Which of the following statements is **not** true of carbon dioxide ?

( )
(a) carbon dioxide is a colourless gas
(b) carbon dioxide has a very faint smell
(c) carbon dioxide has a very slight taste
(d) carbon dioxide can be turned into a liquid at ordinary room temperature by the application of pressure
(e) carbon dioxide is lighter than air

(20) Which of the following is **not** a method of producing carbon dioxide ?

( )
(a) burning carbon in air
(b) the action of dilute acids on all carbonates
(c) the heating of sodium carbonate
(d) the breathing of animals
(e) fermentation

(21) Which of the following is **not** true of acids ?

( )
(a) they turn blue red
(b) they are sour to the taste
(c) they are corrosive
(d) they react chemically with carbonates
(e) they all contain oxygen

(22) Which of the following liquids will turn blue litmus solution red ?

( )
(a) lemon juice
(b) soda water
(c) alcohol
(d) a solution of caustic soda
(e) a solution of sodium hydroxide

(23) Which of the following correctly represents three molecules of sulphuric acid ?

( )
(a) $3HSO_4$
(b) $3H_2SO_4$
(c) $3H_3SO_4$
(d) $3H(SO_4)_3$
(e) H_3SO_4

(24) Which of the following represents four molecules of oxygen correctly ?

( )
(a) $4O_2$
(b) $2O_2$
(b) $4O$
(d) O_2+O_2
(e) O_3+O

(25) Which of the following is the correct formula for copper nitrate ?

( )
(a) $Cu_2(NO_3)$
(b) $CuNO_3$
(c) $Cu_2(NO_2)$
(d) $Cu(NO_3)_2$
(e) $Cu(NO_2)$

(26) Bile is secreted by the cells of the :

( ) (a) duodenum (b) rectum
(c) liver (d) stomach
(e) ileum

(27) Which of the following is a stem and not a root ?

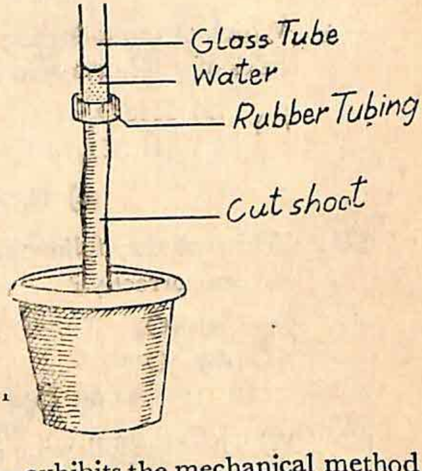
( ) (a) carrot (b) turnip
(c) potato (d) radish
(e) beet

(28) In tropical rain forests, where there is a plentiful supply of water :

( ) (a) cuticular transpiration is almost equal to that of stomatal transpiration
(b) cuticular transpiration is much less than stomatal transpiration
(c) cuticular transpiration is much more than stomatal transpiration
(d) only cuticular transpiration takes place
(e) only stomatal transpiration takes place

(29) What process is demonstrated by the following diagram ?

(a) photosynthesis
(b) transpiration
( ) (c) root pressure
(d) respiration
(e) osmosis



(30) Which of the following exhibits the mechanical method of seed dispersal ?

( ) (a) peas (b) coconut
(c) blackberry (d) cherry
(e) poppy

(31) Which of the following is **not** a drupe ?

( ) (a) plum (b) cherry
(c) date (d) apple
(e) coconut

(32) Which of the following is the cold-blooded animal ?

( ) (a) parrot (b) rabbit
(c) man (d) duck
(e) reptile

(33) Which of the following is not a part of the brain ?

( ) (a) cerebellum (b) pharynx
(c) medulla oblongata (d) cerebral hemisphere
(e) infundibulum

(34) Which of the following is single celled ?

( ) (a) chlamydomonas (b) pandorina
(c) eudorina (d) volvox
(e) spirogyra

(35) Which of the following enzymes emulsifies and decomposes fats ?

( ) (a) rennin (b) diastase
(c) trypsin (d) pepsin
(e) lipase

(36) Animals become rickety and have bad teeth if there is a deficiency of :

( ) (a) vitamin D (b) vitamin A
(c) vitamin E (d) vitamin B
(e) vitamin C

(37) Which of the following substances supplies the greatest amount of protein ?

( ) (a) rice (b) white of egg
(c) vegetables (d) fruits
(e) milk

(38) Blood is carried **away** from the heart by vessels known as :

( ) (a) veins (b) capillaries
(c) arteries (d) trachea
(e) nerves

(39) The subclavians supply blood to :

( ) (a) the head (b) the arms
(c) the hind limbs (d) the intestine
(e) the kidney

(40) Which of the following parts is **not** traversed when air passes from the nose to the lungs ?

( ) (a) pharynx (b) larynx
(c) trachea (d) bronchial tubes
(e) ductless glands

(41) In motor cars the mirror that is used is :

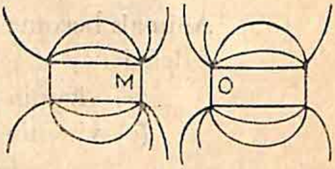
( ) (a) plane (b) concave
(c) combination of plane and concave (d) convex
(e) combination of plane and convex

(42) Which of the following is **not** necessary for photosynthesis ?

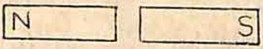
( ) (a) carbon dioxide (b) water
(c) nitrogen (d) heat
(e) light

(43) The following diagram shows the lines of forces between two bar magnets. What is the best deduction that can be made from the facts given ?

(a) M and O are unlike poles
(b) M and O are both N poles
( ) (c) M and O are both S poles
(d) M is N and O is S pole
(e) M is S and O is N pole



(44) A bar magnet was cut into two pieces as shown in the figure. Which of the following statements is true ?



(a) first piece will have only N pole left
(b) second piece will have only S pole left
(c) the magnet is destroyed
(d) both pieces will act as magnets
(e) there will be no poles in either piece

(45) The process by which water is lost from the cells of a

plant by evaporation into the intercellular spaces, and the escape of the vapour through the stomata is known as :

( ) (a) respiration (b) osmosis
(c) imbibition (d) plasmolysis
(e) transpiration

(46) Which of the following is a mixture and **not** a compound ?

( ) (a) water (b) air
(c) iron sulphide (d) sodium chloride
(e) sulphuric acid

(47) Which of the following is **not** an element ?

( ) (a) copper (b) hydrogen
(c) brass (d) sodium
(e) calcium

(48) In the manufacture of ammonia by the synthesis of nitrogen and hydrogen we use ferric oxide as :

( ) (a) reducing agent (b) bleaching agent
(c) oxidising agent (d) dehydrating agent
(e) catalyst

(49) We can say common salt is :

( ) (a) deliquescent (b) hygroscopic
(c) efflorescent (d) hydrated
(e) amorphous

(50) Which of the following is **not** a correct step in testing for soluble sulphates ?

( ) (a) the substance to be tested is dissolved in distilled water (b) the adding of a little sulphuric acid
(c) the adding of a little hydrochloric acid (d) the adding of a solution of barium chloride
(e) to look for a white precipitate

(51) Which of the following hydroxides is very soluble in water ?

( ) (a) ammonium hydroxide (b) calcium hydroxide
(c) barium hydroxide (d) zinc hydroxide
(e) copper hydroxide

PART 3 : DIRECTIONS

In each of the following items certain blanks marked as (A) or (B) are to be filled in. You are to put your answers in the brackets on the left hand side. First item is solved for you as an example.

(1) Plants take in.........A.........and give out.........B...... during the night

(oxygen) A
(carbon dioxide) B

(2) When a solid is dissolved in a liquid the solid is called the......A.........

( ) A

(3) The weight of one c.c. of water is one gram at......A... ....°C

( ) A

(4) The solubility of a substance in a liquid is defined as the weight in grams of the substance which will dissolve in......A......grams of the liquid to form a saturated solution at a given temperature.

( ) A

(5) The poles of a simple magnet are of......A......strength.

( ) A

(6) The oxides which behave either as acidic or as basic oxides are called......A......oxides.

( ) A

(7) The cutting teeth are known as......A......

( ) A

(8) The teeth whose function is tearing flesh are known as ......A......

( ) A

(9) Fish take in......A......for respiration from the water by means of their......B......

( ) A
( ) B

(10) When carbonates of metals are decomposed by heat...A ......is given off and an......B......of the metal is left.

( ) A

( ) B

Directions

Complete and balance the following equations :

(11) H_2SO_4+..............$=Na_2SO_4+$.........

(12) $H_2O+Na=$......$+H_2$

(13) $Cu+2H_2SO_4=CuSO_4+$......$+$...

(14) ....$+4HCl=$.........$+MnCl_2+2H_2O$

(15) $NaCl+H_2SO_4=$.........$+$......

(16) $CaO+HCl=$.........$+$.........

(17) **Directions :**

The following diagram represents half of a flower cut longitudinally. Which pointers indicate the following ? (Write the letter of the pointer in the brackets in front of the item).

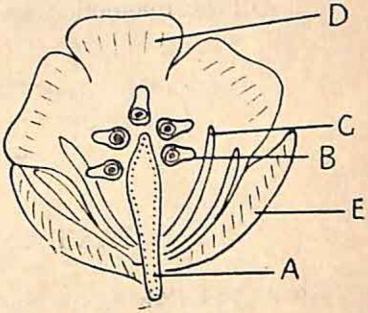
( ) (1) carpel

( ) (2) receptacle

( ) (3) petal

( ) (4) stamen

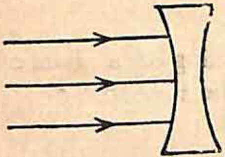
( ) (5) sepal



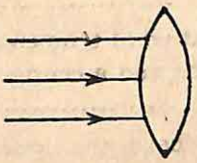
Directions

Each of the following diagrams shows rays of light falling on a solid glass body of the shape shown. Draw the path taken by all the rays as they pass through the glass and out again into the air.

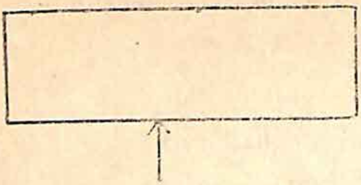
(18)



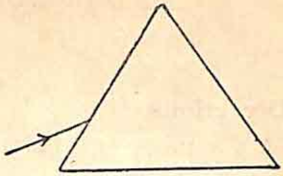
(19)



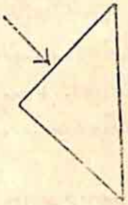
(20)



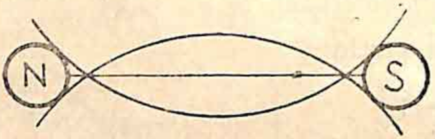
(21)



(22)



(23) In the following diagram N and S are the unlike poles of a magnet. Show by means of arrows the direction of the magnetic lines of force.



PART 4 : DIRECTIONS

Read the following questions and put their answers in the space after the questions. The first item is solved for you as an example :

(1) Which type of balance is used for measuring the mass of a body ?

......physical balance....................

(2) What temperature would 10°C read if converted to Fahrenheit scale ?

....

(3) At what temperature would the reading of a Fahrenheit and a centigrade thermometre be the same ?

....

(4) What is the liquid part of the blood called ?

....

(5) Which method of pollination is most suitable in the case of the rose flower ?

....

(6) A group of about fifty eggs was found. The eggs were yellow in colour, bluntly conical in shape, prettily marked with zig-zag stripes. Name the insect or the animal whose eggs they could be

....

(7) Under adverse conditions—cold, scarcity of food etc.—amœba withdraws all pseudopodia and, rounding itself off, provides itself with a tough covering. What is this state called ?

....

(8) In the following diagram S is the sun, M is the moon and E is the earth.

Which arcs mark off the areas of total eclipse ?

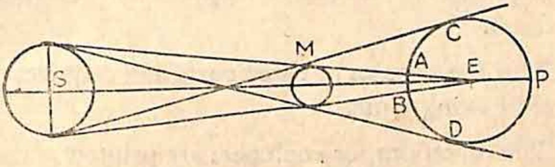
( ) ( )

Which arcs mark off the areas of partial eclipse ?

( ) ( )

Which arcs mark off the areas of no eclipse ?

( ) ( )



Note—Sit up with arms folded when you have done all you can.

(Final Test, English Version)

GENERAL SCIENCE TEST

Please Print

Surname...................................
(in capital letters)

Christian Names.........................
(in capital letters)

Name of School..........................

Your age.........years.........months Your sex...........
(Boy or Girl)

Number of years you have studied general science or science so far.....................................

General Directions

(1) You may use the other side of this sheet for your rough work.

(2) Do not turn this page until the examiner tells you to do so.

(3) This test consists of three parts and requires 40 minutes of working time.

(4) The directions for each part are printed at the beginning of the part. Read them carefully and proceed at once to answer the questions.

(5) Do not spend too much time on any one item. Answer the easier questions first, then return to the harder ones if you have time.

(6) There is no time limit for each part.

(7) You are expected to attempt as many questions as you can from each part.

(8) Guessing will reduce your score, so do not guess.

(9) Use pencil for answering. If you want to change your answers do so by rubbing out and rewriting them.

PART 1: DIRECTIONS

In most of the following sections there are two columns. For each statement or question in the first column there is an answer in the second column. Find the answer and place its letter in the brackets in front of the item in the first column. The first item in Section I is done for you as an example. Answer as many questions as you can in all seven sections in the same way.

Section I

What is the chemical formula for one molecule of :

(L) (1) nitric oxide
( ) (2) ammonium chloride
( ) (3) nitrous oxide
( ) (4) nitric acid
( ) (5) potassium permanganate
( ) (6) sulphuric acid
( ) (7) hydrogen sulphide
(or sulphurated hydrogen)
( ) (8) ammonium sulphate
( ) (9) sodium chloride

(A) H_2SO_4
(B) $KMnO_4$
(C) NaCl
(D) $CaCO_3$
(E) $(NH_4)_2SO_4$
(F) HNO_3
(G) NH_4OH
(H) NH_4Cl
(I) H_2S
(J) N_2O
(K) $Ca(OH)_2$
(L) NO

Section II

What units are the following measured in :

( ) (1) density
( ) (2) power of a lens
( ) (3) latent heat
( ) (4) work
( ) (5) quantity of heat
( ) (6) potential difference
( ) (7) force
( ) (8) power
( ) (9) pressure
( ) (10) electric current

(A) ft. lb.
(B) cm. per sec. per sec.
(C) ampere
(D) volt
(E) dyne
(F) foot pound per sec.
(G) grams per c.c.
(H) calories
(I) dioptre
(J) calories per gram
(K) pound weight
(L) pound weight per. sq. in.

Section III

On the left hand side there are ten processes. On the right hand side there are ten descriptions not in the same order. In the brackets provided write the letter which indicates the statement corresponding to each process in turn.

( ) (1) conduction
( ) (2) electrolysis
( ) (3) convection
( ) (4) refraction
( ) (5) distillation
( ) (6) radiation
( ) (7) evaporation
( ) (8) plasmolysis
( ) (9) osmosis

(A) the method used for separating an insoluble substance from a liquid.
(B) The bending of a ray of light which travels obliquely from one transparent medium to another.
(C) The way in which heat is carried from one point to another in solids.
(D) The method used for separating a dissolved substance from a liquid.
(E) The main way in which heat is carried from one point to another in a liquid.
(F) The collapse and contraction of the protoplasm of a cell due to the passing of water out of the cell.
(G) The process by which liquids and substances in solution tend to pass through permeable, and semi-permeable membrane.
(H) The process of splitting up a substance chemically by means of an electric current.
(I) A process used for obtaining pure water from a solution.
(J) The way in which heat reaches us from any score without appreciably warming the intervening medium.

Section IV

What is the chemical name of each of the common names given in the first column ?

( ) (1) chalk
( ) (2) slaked lime
( ) (3) blue vitriol
( ) (4) washing soda

(A) copper sulphate
(B) sodium chloride
(C) zinc sulphate
(D) calcium carbonate

( ) (5) nitre	(E) potassium hydroxide
( ) (6) common salt	(F) calcium sulphate
	(G) sodium carbonate
	(H) sodium hydroxide
	(I) calcium hydroxide
	(J) sodium nitrate
	(K) zinc carbonate
	(L) potassium nitrate

Section V

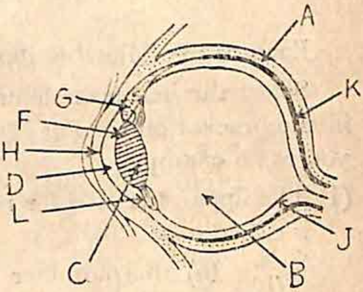
The following diagram shows the section of an eye. Which pointers in the diagram indicate the following parts ?

( ) (1) cornea

( ) (2) pupil

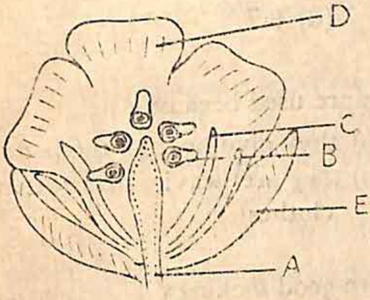
( ) (3) optic nerves

( ) (4) retina



Section VI

The following diagram represents half a flower cut longitudinally. Which pointers indicate the following ? Write the letter of the pointer in the brackets in front of the item.



( ) (1) carpel

( ) (2) receptacle

( ) (3) petal

( ) (4) stamen

( ) (5) sepal

Section VII

The following diagram represents an electric circuit. Which pointer indicates the following :

( ) (1) voltmeter

( ) (2) ammeter

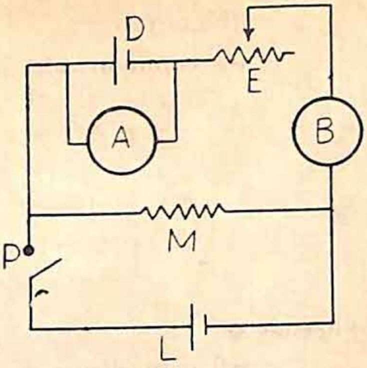
( ) (3) variable resistance

( ) (4) battery

( ) (5) fixed resistance

( ) (6) key

( ) (7) condenser



PART 2 : DIRECTIONS

Each of the following items is followed by five answers.

Select the best possible answer in each case and put its letter in the bracket on the left of the item. Item (1) is solved for you as an example.

(1) The apparatus used for measuring the atmospheric pressure is called a—

(b) (a) thermometer (b) barometer
(c) pyrometer (d) lactometer
(e) hydrometer

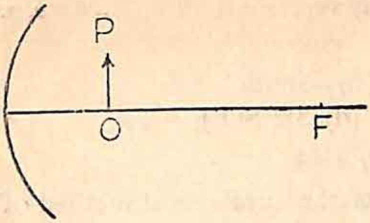
(2) The weight of a piece of iron in air is 10 lbs. and its apparent weight in water is 8·7 lbs. Its specific gravity is—

( ) (a) ·13 (b) 1·87
(c) 7·7 (d) 8·7
(e) 10

(3) In cold climates woollen clothes are used because :

( ) (a) they are durable (b) they give heat to the body
(c) they diminish the loss of heat from the body (d) they are heavier than cotton clothes
(e) they are good looking

(4) The diagram shows an object OP in front of a concave mirror between the principal focus (F) and the mirror. The image formed will be—



( ) (a) the same size as object, real and inverted
(b) magnified, erect, real
(c) magnified, inverted, real
(d) magnified, erect, virtual
(e) smaller than object, inverted, real.

(5) It is **not** true that the image formed by a convex mirror is always :

( )
(a) virtual
(b) diminished
(c) erect
(d) the same distance behind the mirror as the object is in front
(e) not at the same distance behind the mirror as the object is in front.

(6) Which of the following statement is **not** true of carbon dioxide ?

( )
(a) carbon dioxide is a colourless gas
(b) carbon dioxide has a very faint smell
(c) carbon dioxide has a very slight taste
(d) carbon dioxide can be turned into liquid at ordinary room temperature by the application of pressure.
(e) carbon dioxide is lighter than air

(7) Which of the following is not a method of producing carbon dioxide ?

( )
(a) burning carbon in air
(b) the action of dilute acids on all carbonates
(c) the heating of sodium carbonate
(d) the breathing of animals
(e) fermentation

(8) Which of the following liquids will turn blue litmus solution red ?

( )
(a) lemon juice
(b) soda water
(c) alcohol
(d) a solution of caustic soda.
(e) a solution of sodium hydroxide

(9) Which of the following correctly represents three molecules of sulphuric acid ?

( ) (a) $3HSO_4$ (b) $3H_2SO_4$
(c) $3H_3SO_4$ (d) $3H(SO_4)_3$
(e) H_3SO_4

(10) Which of the following exhibits the mechanical method of seed dispersal ?

( ) (a) peas (b) coconut
(c) blackberry (d) cherry
(e) poppy

(11) Which of the following is the cold-blooded animal ?

( ) (a) parrot (b) rabbit
(c) man (d) duck
(e) reptile

(12) Bile is secreted by the cells of the :

( ) (a) deudenum (b) rectum
(c) liver (d) stomach
(e) ileum

(13) Which of the following enzymes emulsifies and decomposes fats ?

( ) (a) rennin (b) diastase
(c) trypsin (d) pepsin
(e) lipase

(14) Animals become rickety and have bad teeth if there is a deficiency of :

( ) (a) vitamin D (b) vitamin A
(c) vitamin E (d) vitamin B
(e) vitamin C

(15) Which of the following substances supplies the greatest amount of protein ?

( ) (a) rice (b) white of egg
(c) vegetables (d) fruits
(e) milk

(16) Blood is carried **away** from the heart by vessels known as :

( ) (a) veins (b) capillaries
(c) arteries (d) trachea
(e) nerves

(17) Which of the following parts is **not** traversed when air passes from the nose to the lungs ?

( ) (a) pharynx (b) larynx
(c) trachea (d) bronchial tubes
(e) ductless glands

(18) In motor cars the mirror that is used is :

( ) (a) plane (b) concave
(c) combination of plane and concave (d) convex
(e) combination of plane and convex

(19) Which of the following is **not** necessary for photosynthesis ?

( ) (a) carbon dioxide (b) water
(c) nitrogen (d) heat
(e) light

(20) The process by which water is lost from the cells of a plant by evaporation into the intercellular spaces, and the escape of the vapour through the stomata is known as :

( ) (a) repiration (b) osmosis
(c) imbibition (d) plasmolysis
(e) transpiration

(21) Which of the following is a mixture and **not** a compound ?

( ) (a) water (b) air
(c) iron sulphide (d) sodium chloride
(e) sulphuric acid

(22) Which of the following is **not** an element ?

( ) (a) copper (b) hydrogen
(c) brass (d) sodium
(e) calcium

(23) In the manufacture of ammonia by the synthesis of nitrogen and hydrogen we use ferric oxide as a :

( ) (a) reducing agent (b) bleaching agent
(c) oxidising agent (d) dehydrating agent
(e) catalyst

(24) We can say common salt is :

( ) (a deliquescent (b) hygroscopic
(c) efflorescent (d) hydrated
(e) amorphous

(25) Which of the following hydroxides is very soluble in water ?

( ) (a) ammonium hydroxide (b) calcium hydroxide
(c) barium hydroxide (d) zinc hydroxide
(e) copper hydroxide

PART 3 : DIRECTIONS

In each of the following items certain blanks marked at (A) or (B) are to be filled in. You are to put your answers in the brackets on the left hand side. First item is solved for you as an example :

(1) Plants take in......A......and give out....B....during the night.

(oxygen) A
(carbon dioxide) B

(2) The teeth whose function is tearing flesh are known a ......A.

( ) A

(3) Fish take in.......A......for respiration from the water by means of their......B......

( ) A
( ) B

(4) When carbonates of metals are decomposed by heat...A ...is given off and an...B...of the metal is left.

( ) A
( ) B

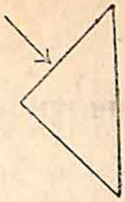
(5) The temperature 10°C, if converted to Fahrenheit scale would be...A...

( ) A

(6) The liquid part of the blood is called...A...

( ) A

(7) The side diagram shows a ray of light falling on a solid glass body of the shape shown. Draw the path taken by the ray as it passes through the glass and out again into air.



(8) In the following diagram S is the sun, M is the moon and E is the Earth.

which arcs mark off the areas of total eclipse ?

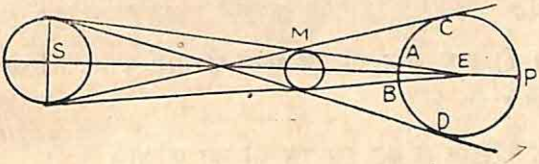
( ) ( )

which arcs mark off the areas of partial eclipse ?

( ) ( )

which arcs mark off the areas of no eclipse ?

( ) ( )



Note—Sit up with arms folded when you have done all you can.

साधारण विज्ञान परीक्षा

निम्न स्थान भरिये

नाम..

(स्पष्ट लिखिये)...

विद्यालय का नाम..

तुम्हारी आयु ....वर्ष........मास......तुम्हारा लिंग..............

(लड़का या लड़की)..........

साधारण विज्ञान या विज्ञान, तुमने कितने वर्ष पढ़ा है

वर्षों की संख्या............

साधारण आदेश

१—इस पत्र की बाईं ओर को तुम अपने अभ्यास कार्य के लिये प्रयोग कर सकते हो।

२—जब तक परीक्षक न कहे इस पत्र को मत खोलो।

३—इस पत्र के तीन भाग हैं और इसे करने का समय ४० मिनट है।

४—प्रत्येक भाग के आरम्भ में, उसे करने के लिये आवश्यक आदेश दे दिये गये हैं। उन्हें ध्यान से पढ़ो और तुरन्त ही कार्य आरम्भ करो।

५—किसी भी प्रश्न पर अधिक समय न लगाओ। पहले पहल सरल प्रश्नों को कर डालो। यदि समय शेष रहे तो कठिन प्रश्नों को करो।

६—प्रत्येक भाग के लिये कोई निश्चित समय नहीं है।

७—तुम से यह आशा है कि प्रत्येक भाग के जितने अधिक से अधिक प्रश्न तुम कर सकते हो करोगे।

८—उत्तरों को अन्दाज से मत लिखो क्योंकि अन्दाज से तुम्हारे प्राप्तांक कम हो जावेंगे।

९—पेंसिल से ही लिखो। यदि किसी उत्तर को बदलना हो तो उसे रबड़ से मिटा कर साफ-साफ लिखो।

प्रथम भाग

नीचे के अधिकांश विभाग दो स्तम्भों में बटे हुए हैं। प्रथम स्तम्भ के प्रत्येक वाक्य का उत्तर दूसरे स्तम्भ में दिया गया है। तुम्हें सही उत्तर ढूँढना है। इस उत्तर के कोष्टक में लिखे हुए अक्षर को प्रथम स्तम्भ के बाईं ओर सामने वाले खाली कोष्टकों में लिखना है। उदाहरण के लिये पहले प्रश्न को हल कर दिया गया है। इसी प्रकार बाकी सात विभागों के प्रश्नों को जहाँ तक तुम कर सकते हो करो।

प्रथम विभाग

नीचे लिखे पदार्थों के एक अणु के लिए कौन सा रसायनिक सूत्र है ?

(अ:) १—नाइट्रिक आक्साइड
( ) २—अमोनियम क्लोराइड
( ) ३—नाइट्रस आक्साइड
( ) ४—शोरे का अम्ल या नाइट्रिक एसिड
( ) ५—पोटास
( ) ६—गन्धक का अम्ल या सल्फ्यूरिक एसिड
( ) ७—हाइड्रोजन सल्फाइड
(सल्फ्यूरेटेड हाइड्रोजन)
( ) ८—अमोनियम सल्फेट
( ) ९—सोडियम क्लोराइड

(अ) $\rightarrow H_2SO_4$
(आ) $\rightarrow KMnO_4$
(इ) $\rightarrow NaCL$
(ई) $\rightarrow CaCO_3$
(उ) $\rightarrow (NH_4)_2SO_4$
(ऊ) $\rightarrow HNO_3$
(ए) $\rightarrow NH_4OH$
(ऐ) $\rightarrow NH_4CL$
(ओ) $\rightarrow H_2S$
(औ) $\rightarrow N_2O$
(अं) $\rightarrow Ca(OH)_2$
(अ:) $\rightarrow NO$

द्वितीय विभाग

निम्नलिखित को किन इकाइयों में नापते हैं ?

( ) १—घनत्व
( ) २—ताल की शक्ति
( ) ३—गुप्त ताप
( ) ४—कार्य
( ) ५—ताप की मात्रा
( ) ६—विभव अन्तर
( ) ७—बल
( ) ८—शक्ति
( ) ९—दबाव
( ) १०—विद्युत धारा

(क) फुट पौंड
(ख) सेन्टीमीटर प्रति सेकिण्ड ·प्रति सेकिण्ड
(ग) एम्पीयर
(घ) वोल्ट
(ङ) डाइन
(च) फुट पौंड प्रति सेकिण्ड
(छ) ग्राम प्रति घ० सेन्टीमीटर
(ज) कैलोरीज
(झ) डाइप्टर (Dioptre)
(ञ) कलोरीज प्रति ग्राम
(ट) पौण्ड भार
(ठ) पौण्ड भार प्रति वर्ग इंच

तृतीय विभाग

नीचे बाईं ओर नौ विधियाँ दी हुई हैं। दाहिनी ओर उन्हीं के दस वर्णन दिए हुए हैं। ये वर्णन क्रम से नहीं हैं। विधियों के उपयुक्त वर्णन ढूँढ़कर उनके कोष्टकों के अक्षरों को बाईं ओर दिए गए खाली कोष्टकों में भरो।

( ) १—संचालन

( ) २—विद्युत विश्लेषण

( ) ३—सम्वहन

( ) ४—आवर्तन

( ) ५—श्रावन

( ) ६—विकिरण

( ) ७—वाष्पीकरण

( ) ८—रसकोच

( ) ९—रसाकर्षण

(अ) एक अनघुल पदार्थ को किसी द्रव से अलग करने की विधि।

(आ) एक पारदर्शक माध्यम से दूसरे पारदर्शक माध्यम को तिरछी जाने वाली प्रकाश को किरण का झुक जाना।

(इ) ठोसों में एक बिन्दु से दूसरे तक ताप जाने की विधि।

(ई) घुले हुए पदार्थ को द्रव से अलग करने की विधि।

(उ) द्रव में ताप का एक बिन्दु से दूसरे बिन्दु तक जाने की मुख्य विधि।

(ऊ) कोष से पानी निकल जाने पर उसके प्रोटोप्लाज्म का सिकुड़ना और बैठ जाना।

(ए) घोल में स्थित द्रवों तथा पदार्थों का शोषक तथा अर्धशोषक झिल्लियों से होकर प्रवेश करने की विधि।

(ऐ) विद्युतधारा द्वारा किसी पदार्थ का रसायनिक विश्लेषण करने की विधि।

(ओ) किसी घोल से शुद्ध जल प्राप्त करने की विधि।

(औ) बीच के माध्यम को बिना अच्छी तरह गर्म किये हुए ताप का किसी स्थान से हम तक पहुँचने की विधि।

चतुर्थ विभाग

नीचे दिये हुये प्रथम स्तम्भ को सामान्य संज्ञाओं के रसायनिक नाम बताओ—

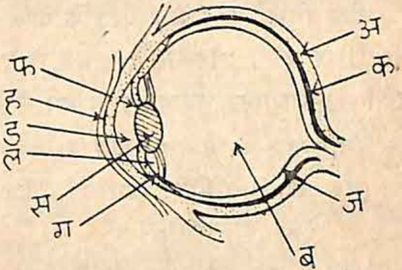
( ) १—खड़िया
( ) २—बुझा चूना
( ) ३—नीला थोथा
( ) ४—धोने का सोडा
( ) ५—शोरा
( ) ६—साधारण नमक

(क) कॉपर सल्फेट
(ख) सोडियम क्लोराइड
(ग) जिन्क सल्फेट
(घ) कैल्शियम कार्बोनेट
(ङ) पोटैशियम हाइड्रा ऑक्साइड
(च) कैल्सियम सल्फेट
(छ) सोडियम कार्बोनेट
(ज) सोडियम हाइड्रोऑक्साइड
(झ) कैल्सियम हाइड्रोऑक्साइड
(ञ) सोडियम नाइट्रेट
(ट) जिन्क कार्बोनेट
(ठ) पोटैशियम नाइट्रेट

पंचम विभाग

नीचे दिया गया रेखाचित्र आँख की बनावट दिखलाता है। रेखाचित्र में कौन सा तीर सूचक निशान नीचे लिखे भागों को सूचित करता है ?

( ) १—कार्निया
( ) २—तारा
( ) ३—दृष्टि नाड़ियाँ
( ) ४—मूर्ति-पट

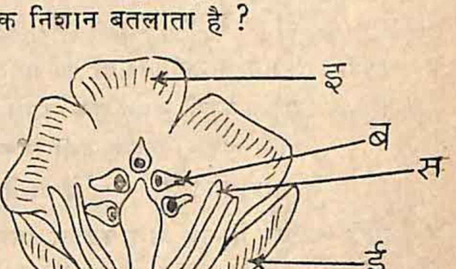


षष्टम विभाग

नीचे दिया गया रेखाचित्र लम्बाई में कटा हुआ फूल का आधा भाग है। निम्नलिखित को कौन-सा तीर सूचक निशान बतलाता है ?

(वस्तुओं के सामने के कोष्टों में तीर सूचकों के अक्षरों को लिखो)

( ) १—अन्डयों
( ) २—स्तम्भक
( ) ३—पंखुडी
( ) ४—पुंकेसर
( ) ५—अंकुडी



सप्तम विभाग

नीचे दिये रेखाचित्र में विद्युत परिपथ दिखाया गया है। नीचे लिखी वस्तुओं को कौन-सा तीर सूचक बतलाता है ?

( ) १—वोल्ट मापी

( ) २—अम्मापी

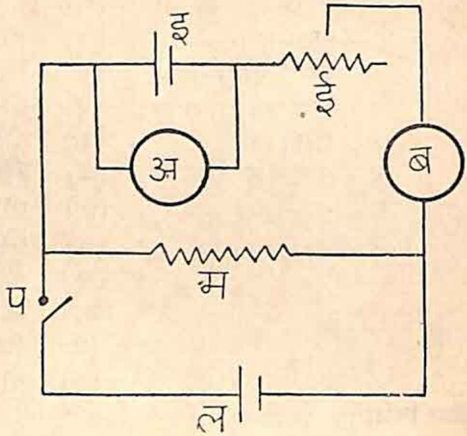
( ) ३—परिवर्ती प्रतिरोधक

( ) ४—बैटरी

( ) ५—निश्चित प्रतिरोधक

( ) ६—कुंजी

( ) ७—संधारित्र
(Condenser)



द्वितीय भाग

नीचे लिखे हुए प्रत्येक प्रश्न के पाँच उत्तर दिये गये हैं। प्रत्येक प्रश्न से सही उत्तर छाँटो और उसके सामने लिखे अक्षर को कोष्टक में बायीं ओर लिखो। प्रथम प्रश्न उदाहरण के रूप में हल किया गया है।

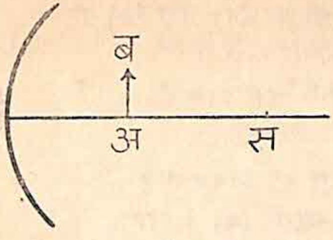
१—वायुमण्डल का दबाव नापने वाले यन्त्र को कहते हैं—
(ब) (अ) थर्मामोटर (ब) बैरोमीटर (स) पाइरोमीटर (द) लैक्टोमोटर (य) हाइड्रोमीटर।

२—लोहे के एक टुकड़े का हवा में भार १० पौण्ड है और उसका पानी में भार ८·७ पौण्ड है उसका आपेक्षिक घनत्व बराबर होगा।
( ) (अ) ·१३ (ब) १·८७ (स) ७·७ (द) ८·७ (य) १०

३—ठण्डी जलवायु में ऊनी कपड़ों का प्रयोग किया जाता है क्योंकि :—
( ) (अ) वे काफी समय तक चलते हैं। (ब) वे शरीर को गर्मी देते हैं। (स) वे शरीर से कम गर्मी निकलने देते हैं। (द) वे सूती कपड़ों से भारी होते हैं। (य) वे देखने में सुन्दर लगते हैं।

४—नीचे के चित्र में एक वस्तु अवतल दर्पण के सामने मुख्य संगम (स) तथा दर्पण के बीच रक्खी है उसका बिम्ब :—

( ) (अ) वस्तु के बराबर, वास्तविक और उल्टा होगा। (ब) वस्तु



से बड़ा, सीधा और वास्तविक होगा। (स) वस्तु से बड़ा, उलटा और वास्तविक होगा। (द) वस्तु से बड़ा, सीधा और काल्पनिक होगा। (य) वस्तु से छोटा उल्टा और वास्तविक होगा।

५—यह सत्य नहीं है कि उत्तल दर्पण से बना हुआ बिम्ब सदैव :—

( ) (अ) काल्पनिक होगा। (ब) मन्द होगा। (स) सीधा होगा। (द) दर्पण से उतनी ही दूर पीछे होगा जितनी कि वस्तु दर्पण के सामने दूर है। (य) दर्पण के पीछे उतनी दूरी पर नहीं होगा जितनी कि वस्तु दर्पण के सामने है।

६—इसमें कौन सा कथन कार्बन डाई ऑक्साइड के लिये सत्य नहीं है—

( ) (अ) कार्बन डाई ऑक्साइड रंगहीन गैस है। (ब) कार्बन डाई ऑक्साइड में हल्की गन्ध होती है। (स) कार्बन डाई ऑक्साइड में हल्का स्वाद होता है। (द) कार्बन डाई ऑक्साइड साधारण कमरे के तापक्रम में दबाव द्वारा द्रवीभूत की जा सकती है। (य) कार्बन डाई ऑक्साइड हवा से हल्की है।

७—निम्न में से कौन-सा ढङ्ग कार्बन डाई ऑक्साइड बनाने का नहीं है ?

( ) (अ) कार्बन को हवा में जलाने से। (ब) हल्के अम्ल की भिन्न-भिन्न कार्बोनेटों पर क्रिया से। (स) सोडियस कार्बोनेट को गरम करने से। (द) जानवरों के साँस लेने से। (ध) किन्वन, सड़ने की विधि से (Fermentation)

८—नीचे लिखे द्रवों में से कौन-सा नीले लिटमस को लाल करता है :—

( ) (अ) नींबू का रस। (ब) सोडा वाटर। (स) अल्कोहल। (द) कास्टिक सोडा का घोल। (ध) सोडियम हाईड्रो ऑक्साइड का घोल।

९—निम्नलिखित में कौन-सा सूत्र गन्धक के अम्ल के तीन अणु सम्बोधित करता है—

( ) (अ) $3HSO_4$ (ब) $3H_2SO_4$ (स) $3H_3SO_4$ (द) $3H(SO_4)_3$ (ध) H_3SO_4.

१०—निम्नलिखित में कौन से बीजों में यांत्रिक विधि के द्वारा विकिरण होता है—

( ) (अ) मटर (ब) नारियल (स) काली वेर (द) चैरी (ध) पौंपी ।

११—निम्नलिखित में कौन ठण्डे खून का जानवर है—

( ) (अ) तोता (ब) खरगोश (स) मनुष्य (द) बतख
(ध) रैप्टाइल (Reptile)

१२—निम्नलिखित में से किसके कोष्टकों से पित्त रस निकलता है—

( ) (अ) ड्यूडेनम (ब) रेक्टम (स) यकृत (द) आमाशय
(ध) क्षुद्र आंत (Ileum).

१३—निम्नलिखित में कौन सा विकर (Enzyme) वसा को प्रतिलिम्बित करता और गलाता है—

( ) (अ) रेनिन (ब) डाईस्टेज (स) ट्रिपसिन (द) पेपसिन
(ध) लाइपेज ।

१४—निम्नलिखित में किस की कमी के कारण जानवरों के खराब दाँत तथा कमजोर हड्डियाँ हो जाती हैं—

( ) (अ) विटामिन 'डी' (ब विटामिन 'ए' (स) विटामिन 'ई'
(द) विटामिन 'बी' (ध) विटामिन 'सी' ।

१५—निम्नलिखित में किस पदार्थ में प्रोटीन की सबसे अधिक मात्रा प्राप्त होती है—

( ) (अ) चावल (ब) अण्डे की सफेदी (स) सब्जी (द) फल (ध) दूध ।

१६—हृदय से रुधिर को ले जाने वाली नालियों को—

( ) (अ) शिरायें (ब) कोशिकायें (स) धमनियाँ (द) ट्रँकियाँ
(ध) नाड़ियाँ कहते हैं ।

१७—जब वायु नाक से होकर फेफड़ों में पहुँचती है तो नीचे लिखे हुए में से किस भाग से होकर नहीं जाती—

( ) (अ) ग्रस्वनिका (ब) स्वर यन्त्र (स) टेंटुआ (द) वायु प्रणालियाँ
(ध) प्रणाली विहीन ग्रन्थियाँ ।

१८—मोटरकारों में किस दर्पण का प्रयोग होता है—

( ) (अ) समतल दर्पण (ब) अवतल (स) समतल और अवतल मिला हुआ (द) उत्तल (ध) समतल और उत्तल मिला हुआ ।

१९—निम्नलिखित में से कौन फोटोसिन्थेसिस के लिये आवश्यक नहीं है—

( ) (अ) कार्बन डाई आक्साइड (ब) पानी (स) नाइट्रोजन (द) गर्मी
(ध) प्रकाश ।

२०—वाष्पीकरण द्वारा जल का पौधे के कोष्टकों से कोष्टकों के बीच के भाग में जाने की विधि तथा स्टोमेटा से भाप निकलने की विधि को कहते हैं—

( ) (अ) श्वासोच्छ्वास क्रिया (ब) आसमोसिस (रसार्षण) (स) अन्तचूषण (द) रसकोच (ध) ट्रांसपाइरेशन।

२१—इनमें कौन-सा मिश्रण है यौगिक नहीं—

( ) (अ) जल (ब) हवा (स) लोहे का सल्फाइड (द) नमक (ध) गन्धक का अम्ल।

२२—इनमें कौन-सा तत्व नहीं है :—

( ) (अ) तांबा (ब) हाइड्रोजन (स) पीतल (द) सोडियम (ध) कैलसियम।

२३—नाइट्रोजन और हाइड्रोजन के संश्लेषण द्वारा अमोनिया को व्यापारिक विधि से बनाने में फैरिक आक्साइड का प्रयोग किस रूप में किया जाता है ?

( ) (अ) अवकारक (ब) विरंजन कारक (स) आक्सीकारक (द) निर्जलीकारक (ध) उत्प्रेरक।

२४—हम साधारण रूप से सामान्य नमक को कह सकते हैं—

( ) (अ) प्रक्लेद्य (ब) आर्द्रताग्राही (स) प्रस्फुटित (द) जलयोजित (ध) अकेलास।

२५—इनमें कौन-सा हाइड्रो आक्साइड पानी में घुलनशील है—

( ) (अ) अमोनियम हाइड्रो आक्साइड (ब) कैलसियम हाइड्रो आक्साइड (स) बेरियम हाइड्रो आक्साइड (द) जस्ते का हाइड्रो आक्साइड (ध) तांबे का हाइड्रो आक्साइड।

भाग ३

आदेश :—नीचे लिखे प्रश्नों में 'अ' और 'ब' रिक्त स्थानों की पूर्ति करनी है। तुमको अपने उत्तर कोष्टकों के अन्दर लिखने हैं। प्रथम प्रश्न उदाहरणार्थ तुम्हारे लिये हल किया गया है।

१—पौधे रात्रि के समय (अ)·······लेते हैं और (ब)········देते हैं।

(आक्सीजन) (अ) (कार्बन डाइ आक्साइड) (ब)

२—दाँत जिनका कार्य मांस चीरना है उन्हें (अ)········कहते हैं। ( ) (अ)

३—मछलियाँ श्वासोच्छ्वास क्रिया के लिये अपने ……(अ) द्वारा पानीसे (ब) ……लेती हैं।

( ) (अ) ( ) (ब)

४—जब धातु के कार्बोनेट्स का गर्मी के द्वारा विच्छेदन किया जाता है तो……(अ) निकलता है और धातु का……(ब) बच रहता है।

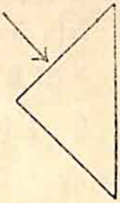
( ) (अ) ( ) (ब)

५—१०० सेन्टीग्रेट तापक्रम को फाइरनहाइट स्केल में बदलने से……(अ) होता है।

( ) (अ)

६—रक्त के द्रव वाले भाग को……(अ) कहते हैं। ( ) (अ)

७—सामने एक ठोस काँच का चित्र खिंचा है इस पर एक प्रकाश की किरण पड़ रही है। यह किरण कांच के अन्दर तथा कांच से बाहर आकर किस तरह चलेगी। चित्र में अंकित करो।



८—नीचे दिये गये चित्र में 'स' सूर्य, 'म' चन्द्रमा और 'इ' पृथ्वी को अङ्कित करता है :—

कौन-से चाप पूर्ण ग्रहण के भाग को संबोधित करते हैं ?

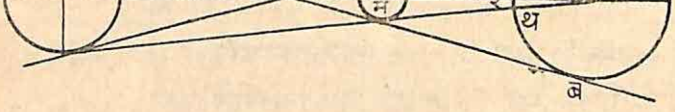
( ) ( )

कौन-से चाप अर्द्ध-ग्रहण के भाग को सम्बोधित करते हैं ?

( ) ( )

कौन-से चाप बिना ग्रहण वाले भाग को सम्बोधित करते हैं ?

( ) ( )



नोट :—कार्य समाप्त करने पर अपने स्थान पर बैठे रहो।

CHAPTER 6

DIAGNOSTIC TESTS AND REMEDIAL TEACHING

What is a Diagnostic Test?

A DIAGNOSTIC TEST IS DESIGNED TO REVEAL A PERSON'S strength and weakness in one or more areas of the field being tested. It assists the teacher in determining exactly where the learning or teaching has been successful and where it failed. This test is also a school subject test like the general achievement test but its purpose is quite different. Its purpose is to inquire into the nature of the difficulties which a weak pupil is having with his school subjects so that proper remedial instructions may be applied. In such tests a single, total or composite score is of little or no significance, but on which the part scores or the percentages of correct responses to indívidual items are the measures sought. Test may be diagnostic in various degrees. A test in English correctness, for example, may break the whole field up into such divisions as spelling, capitalization, punctuation, grammar, and usage, yielding a part score for each division, or may still further analyze each of these divisions, splitting the section on punctuation into tests of use of the comma, period and semicolon etc. Or it may make an even detailed analysis, considering separately, for example, each of the types of situations in which, for instance, the comma may be used. To the degree, then, that the emphasis in the test is placed upon part scores or upon percentages of responses to individual items, that test is of the diagnostic type. To the degree that the emphasis is placed upon a single total score, designed to yield a measure of general achievement, the test is of the general achievement type.

Teachers who have a little or no knowledge of diagnostic test spend much of their time in teaching over and over again when they see that their pupils have not followed them. For an illustration if the pupils find it difficult to compute the problems on 'ratio and proportion' it does not mean that they do not know the method of computing these problems. These problems require a thorough knowledge of multiplication, division or cross multiplication etc. It is just possible that some pupils may be weak in multiplication and some in division. If these pupils be given the thorough instructions in multiplication and division according to their need it will be easy for them to solve the steps in the problems of ratio and proportion. It is of no use wasting the time to repeat the lesson over and over again. The weak point of the pupils should be diagnosed and more attention should be given to remove that difficulty. Many teachers complain that their pupils are weak in civics or history. The real difficulty may be the poor reading of the pupils and hence they fail to organize their thoughts properly. Teachers not knowing the nature of the difficulty may not be able to find out the proper remedy to correct their pupils. A wise teacher will give more practice to his pupils in reading for such cases.

If a pupil is weak in spelling, it will not correct him to write a long paragraph which the teacher dictates each day. That paragraph may contain hundred and one other words which the pupil can spell well. It is only the waste of time to give practice in spelling of such words which the pupil knows well. The analysis of pupil's written work can reveal those words in which more practice for spelling is needed. More practice should be given in spelling of such words only. Again the teaching of phonetics may prove helpful. Such written work should be given to the pupils in which these words reappear again and again. What we mean is that our teaching should not be such that may scatter its efforts on a wide field without sharply defined aims.

In the beginning of any lesson we should test the previous knowledge of the pupils. If a boy joins the high-school after the primary school we should test him in all that he has read in the primary stage. Suppose we are to begin a lesson on 'Ratio and Proportions' in the High School. It is just possible that the boy might not be able to pick up the steps in division and multiplication in primary school from where he comes. If we start teaching this new topic without removing his minor difficulties in multiplication and division he will not follow it. The teacher should not take the attitude as 'this multiplication and division is the job of primary school teacher, why should I teach it here !' Teacher is to teach according to child centre approach.

If in class eighth we are to teach the geometry we should prepare a comprehensive test to examine the previous knowledge of the pupils as regards the angles, triangles, lines, congruency of triangles, areas of triangles and rectangles etc. By analysing this test we can discriminate between the weak and strong pupils. We shall also be able to find out the weak point of pupils. If we see most of the pupils are weak in 'area of the triangles' more attention should be given to it. We may inform the individuals about their mistakes. Special home-work emphasizing the drill sums on the weak points should be given. Ten minutes of each period should be devoted for discussing the assignments. This method may correct the pupils in a month or two and may save their eighth class career.

A programme for remedial teaching

Any diagnostic procedure must be based on extensive research to determine what types of errors are made. Once the error are tested, it is necessary to devise test procedures to reveal clearly which errors the pupil makes. After diagnosing the weakness or the mistakes of the pupil the teacher needs to give him remedial teaching. If a doctor diagnoses that the patient is suffering from cholera and does not give any remedy for its cure it will have no value at all. Diagnosis unless supplemented by remedy is worthless not only in medicine but also in the educational field. Much work has been done so far in diagnostic testing and remedial teaching for reading, arithmetic and other school subjects.

Diagnostic teaching and Remedial teaching in reading

Teachers in the elementary and secondary schools find that they are perpetually faced with the problem of what to do with the child who is deficient in reading. Remedial instructions in reading should be a normal part of the reading programme at all grade levels. Researches have proved that retardation in reading may seriously affect the personality development of pupils. Pupils desire recognition. When this basic craving is frustrated, certain types of escape mechanism often make their appearance. Juvenile delinquency and all sorts of anti-social conduct can be traced to the humiliating effects caused by the failure of pupils to make adequate progress in reading.

A first step in any remedial reading programme is to identify those pupils most in need of attention. Procedures which can be employed for this purpose include observing pupils while they study ; studying the eye movements during reading ; using interest inventories ; graded sets of books ; standardised silent reading test, standardized oral reading tests, and informal teacher—made tests of reading. Schools more frequently used standardized silent reading tests for locating their poor readers than any other method.

Pupils selected for remedial work should never be made to feel that they are different or inferior to other pupils. The special reading classes should not be labeled remedial.

A careful study should be made of each pupil who is selected for remedial work in reading in an effort to locate the cause of his difficulty. Pupils fail to learn to read for a variety of reasons. Some of the factors that may lie at the bottom of a given pupils special inability in reading are : poor physical condition, sensory defects, unsatifactory home conditions, low intelligence, left-handedness, emotional tension, lack of interest, meager background of experience, and lack of reading practice. One of the chief and most obvious of these causes is lack of reading practice. A sure care for retardation in reading is to get the pupil to read. But why is it, one pupil reads extensively and another avoids reading entirely ? A careful investigation may reveal that physical or emotional factors, home conditions, and lack of interest are responsible for it. The diagnosis should, therefore, attempt to discover why a pupil has no interest in reading. It should also try to discover any basic non-reading interests the pupil may possess which may be used as the foundation for developing reading interests.

The specific nature of the remedial treatment to be given to a pupil should depend upon the particular diagnosis which is made of his case. No two pupils are exactly alike either with respect to the difficulties they may have or with respect to the interests and goals they possess. Despite this fact, there are several rules of procedure which have wide applicability. Five of these are : (1) begin where the pupil is, (2) inform him frequently by means of charts and graphs of the progress he is making, (3) see that the exercises engaged in satisfying some basic goal of the pupil, (4) frequently commend the pupil for work well done, and (5) supply a variety of exercises and activities so that the work will not become monotonous.

A pupil's reading vocabulary can be extended through extensive reading ; the use of an appropriate dictionary ; systematic word study ; the study of prefixes, suffixes, and the root from which words are derived. Wide reading method is most useful. Power of comprehension in reading can be increased through vocabulary enrichment, the broadening of experience backgrounds, and through specific practice in reading materials of a variety of types.

Diagnostic testing and remedial teaching in Arithmetic

Many pupils in the elementary schools and high schools are found who are deficient with respect to the most elementary arithmetical processes. Defective preparation in arithmetic has been shown by numerous pupils who have accepted positions in industry. There is a felt need for remedial teaching in arithmetic.

Evidence is ample that pupils of average or even below average mental ability can greatly increase their arithmetical skills when suitable remedial measures are instituted. Cases are on record of pupils who have improved as much as four grades as the result of one terms work. Entire classes have been brought up as much as two grades in four or five months.

Diagnostic tests in arithmetic are of great value in identifying specific weaknesses which pupils possess and in pointing the way toward appropriate remedial instruction.

Survey tests in arithmetic are useful in giving the grade level at which a pupil performs, or in furnishing information as to which of the fundamental processes (addition, subtraction, division, multiplication) causes him the most trouble. Such a test, for example, might show that a pupil is up to par in all processes except division. It would not, however, show which types of division problems he can do and which he cannot do. Since survey tests frequently do not probe deeply enough to show the particular error or errors that may be at the bottom of a pupil's difficulty in arithmetic, it is often highly desirable to use specially constructed diagnostic tests to furnish additional information.

Buswell-John Diagnostic test for fundamental processes in Arithmetic

This is an individual test, which provides the teacher with an insight into the specific errors which a pupil makes in arithmetic. The test consists of two sections—the 'Pupil's Work Sheet' and the 'Teacher's Diagnostic Chart'. The pupil's work sheet consists of a variety of problems in addition, subtraction, multiplication, and division, which the pupil works while thinking aloud. While the pupil proceeds, the teacher uses his diagnostic chart to keep a record of his specific errors and methods of work. The technique of having the pupil 'speak out' while working his problems is very important. Errors can be discovered in this way which would be entirely overlooked by a mere examination of the test after the pupil had completed it. Some pupils, for example, in doing a problem in addition, skip around the column in selecting numbers to be added rather than proceeding directly down the column taking each number as it comes. This defective work habit would be undetected unless the technique used in the Buswell-John test were employed.

The test is so designed as to check a great variety of possible errors. The types of errors generally detected in the case of addition are shown in the following chart :

Sample page of the Buswell-John Teacher's Diagnostic Chart for Individual Difficulties in the Fundamental Processes of Arithmetic.

Teacher's Diagnosis
for pupil......

Teachers Diagnostic Chart
for
Individual difficulties in
Fundamental Processes of
Arithmetic.

Name.............. School.............. Age........
Date of Diagnosis... Class......... I.Q...........
Add....... Subt....... Mult....... Div.......
Teacher's preliminary diagnosis...........

Addition :—(Place a check before each habit observed in the Pupil's work)

1. Errors in combinations.
2. Counting.
3. Added carried number last.
4. Forgot to add carried number.
5. Repeat work after partly done.
6. Added carried number irregularly.
7. Wrote number to be carried.
8. Irregular procedure in column.
9. Carried wrong number.
10. Grouped two or more numbers.
11. Splits numbers into parts.
12. Used wrong fundamental operation.
13. Lost place in columns.
14. Depended on visualization.
15. Disregard column position.
16. Omitted one or more digits.
17. Errors in reading numbers.
18. Dropped one or more tens.
19. Derived unknown combination from familiar one.
20. Disregard one column.
21. Error in writing answer.
21. Carrying when there was nothing to carry.
23. Used scratch paper.
24. Added some digits in two columns.
25. Wrote carried number in answer.

(Write observation note on pupil's work in space opposite example).

(1) 5 6 2 3		(2) 2 8 9 4	
(3) 2 8 9 4		(4) 52 42 13 39	

Complete directions for making a diagnosis of the pupil's work in arithmetic are contained in the manual of directions, which also includes numerous specific examples of each of the types of errors which pupils frequently make. After a teacher has become skilled in the use of this test, it is possible in about 20 minutes time to discover the characteristic work habits and errors of most pupils.

There are several tests for addition, subtraction, division, fraction etc. Below is given the true copy of a diagnostic Arithmetic Test on Addition from 'Diagnostic and Attainment Testing' by F. J. Schonell and F. E. Schonell.

Directions

These are all add sums, work them across the page. You will have plenty of time to do these sums.

	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
A	1+1=	0+0=	2+2=	2+1=	1+3=
B	2+0=	3+1=	3+3=	5+5=	4+1=
C	1+6=	4+0=	4+4=	1+7=	6+1=
D	7+1=	1+8=	1+5=	5+0=	7+7=
E	2+5=	3+2=	1+4=	0+2=	6+6=
F	4+3=	8+8=	2+9=	2+8=	5+4=
G	9+2=	9+1=	9+9=	3+0=	0+7=
H	0+1=	6+0=	6+2=	2+4=	0+4=
I	8+2=	6+4=	1+9=	4+5=	0+8=
J	8+3=	5+1=	7+2=	0+3=	1+2=
K	0+5=	4+2=	3+6=	8+1=	0+9=
L	1+0=	5+3=	2+7=	3+5=	7+0=
M	2+3=	5+2=	3+4=	8+0=	2+6=
N	6+3=	4+8=	0+6=	7+3=	8+4=
O	9+0=	3+7=	6+5=	7+4=	4+6=
P	3+8=	5+6=	3+9=	7+6=	9+3=
Q	6+7=	4+7=	8+9=	5+7=	9+4=
R	7+5=	5+9=	4+9=	8+6=	7+8=
S	9+5=	8+7=	6+9=	9+8=	9+7=
T	6+8=	9+6=	8+5=	5+8=	7+9=

Answers to this Test

When pupils are correcting their own results teachers should read the answers across the pages in order in which they have been worked.

	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
A	2	0	4	3	4
B	2	4	6	10	5
C	7	4	8	8	7
D	8	9	6	5	14
E	7	5	5	2	12
F	7	16	11	10	9
G	11	10	18	3	7
H	1	6	8	6	4
I	10	10	10	9	8
J	11	6	9	3	3
K	5	6	9	9	9
L	1	8	9	8	7
M	5	7	7	8	8
N	9	12	6	10	12
O	9	10	11	11	10
P	11	11	12	13	12
Q	13	11	17	12	13
R	12	14	13	14	15
S	14	15	15	17	16
T	14	15	13	13	16

NORMS FOR SCHONELL'S DIAGNOSTIC TEST

Time—3 Minutes

Scores and Equivalent Arithmetic Ages.

Score	Arith. Age		Score	Arith. Age		Score	Arith. Age		Score	Arith. Age	
	Yrs.	Months		Yrs.	Months		Yrs.	Months		Yrs.	Months
25	7	0	43	8	6	64	10	0	82	11	8
27	7	1	44	8	7	65	10	1	83	11	8
28	7	2	45	8	8	66	10	2	84	11	10
29	7	3	46	8	9	67	10	3	85	12	0
30	7	4	47	8	9	68	10	4	86	12	4
31	7	5	48	8	10	69	10	5	87	12	6
32	7	6	49	8	11	70	10	6	88	12	8
33	7	7	50	9	0	71	10	7	89	12	10
34	7	8	51	9	1	72	10	8	90	13	0
35	7	9	52	9	2	73	10	10	91	13	2
36	7	10	53	9	3	74	10	11	92	13	4
37	7	11	54	9	3	75	11	0	93	13	6
38	8	1	55	9	4	76	11	1	94	13	10
39	8	2	56	9	5	77	11	2	95	14	2
40	8	3	57	9	6	78	11	4	96	14	6
41	8	4	58	9	7	79	11	5	97	14	9
42	8	5	59	9	8	80	11	6	98	15	0
			60	9	9	81	11	7			
			61	9	10						
			62	9	11						
			63	10	0						

Only the diagnosis of the weakness of the pupil in arithmetic will not do but the diagnosis should make a thorough appraisal of the pupil's physical condition, his visual and auditory equipment, his achievement in other school subjects, his general mental ability, his home conditions, his attitudes and interests, his work habits, and specific error patterns. Only by taking all possible background and personal factors into consideration as well as specific difficulties in arithmetic is it possible to form an accurate judgment of the cause of a pupil's disability.

This remedial teaching which follows the diagnosis of individual difficulties should, of course be geared to weaknesses that have been found in pupil's processes and habits of work. It should also take into account the needs and goals of the pupil. It should be more than mere drill upon processes and combinations that have not been mastered. Drill or practice is necessary, but this should be highly stimulating and related to the life-career motives of the pupil wherever possible. Pupils should be kept informed of the progress they are making, and new procedures and devices should be introduced into the instructions from time to time to keep monotony from setting in.

At present there is a dearth of diagnostic tests, in school subjects, in India. Remedial teaching cannot go without diagnostic tests and learning cannot be stimulated and made effective without remedial teaching. Hence there is an urgent need for the construction of diagnostic tests in various school subjects.

CHAPTER 7

INTELLIGENCE AND ITS NATURE AND NURTURE

Introduction

FROM THE ANCIENT TIMES PEOPLE BEGAN TO REALIZE THAT some people were intelligent and some were dull. They had hardly the clear conception of the means of ascertaining exactly what this intelligence was or to what extent one man was more intelligent than another. They could estimate that man having a fine 'knowledge' cannot be so intelligent. The word intelligence had no psychological significance. At the end of the last century many attempts were made to define 'intelligence' but all unsuccessful. Gall, Lavator and a few others assessed an individual's intelligence on the basis of the size and shape of his head, his sensitivity to light, sound, and touch, etc. The earliest observation of historical importance was the one which was noted in connection with an astronomical observation in Greenwich Laboratory in 1795, where one observer differed very much from his superior officer in his record of the time of the transit of a star and as a result was accused of negligence of duty and discharged from his post. Later on Cattell could well recognize the significance of these individual differences. As a result of his experiment carried in Wundt's laboratory he was led to believe that individuals varied greatly from one another in motor movements and sensory discrimination. The first man who launched to study intelligence successfully was a French man, Alfred Binet (1857—1911). Binet was the French psychologist. The problem before him was that a few boys in Paris schools could not take the benefit from school education, however better coach they received. On the other hand others had a better capacity to be benefited by school education though they received a comparatively poorer type of teaching. He wanted to distinguish between these two sets of boys. The former he called mentally retarded. Since the mentally retarded learn so much slowly than others much time was wasted, before it was discovered what is beyond the pupil's capacity and what was within his capacity. He started by collecting a number of tests from the most common experiences of children and applied them to about two hundred children in Paris. Then he graded the tests according to their difficulty. Thus appeared his scale in 1905. These tests were further elaborated by him and rearranged in an age scale in his 1908 and 1911 versions as will be discussed in the later part of this chapter.

It can be said that before Binet people did not recognize the significance of intelligence but we cannot deny the fact that people used to take interest in the improvement of methods of diagnosing and treating feeble-minded and backward children. The earliest case of a feeble-minded person properly studied was that of a wild boy who was found by some hunters in 1797 in the woods of Aveyron in Southern France. This wild boy of Aveyron was brought to Paris and aroused the keenest interest among the psychologists of the time.

They set about training the boy in the most scientific manner. Itrad, the famous psychologist of feeble-minded, tried that case. After a hard labour when he could not bring him to a level of a normal child he gave up the work. After Itrad his pupil Seguin started the work of coaching the feeble-minded. Soon psychologists began to realize that these feeble-minded children were not a class by themselves, but were only at the lower end of the general population arranged in ascending order according to their mental abilities.

Nature of Intelligence

Now let us see what intelligence means.

The question 'what is intelligence' is a theme of debate. Different psychologists define intelligence in a different way, but we can easily group them in three categories :

(1) Those that regard intelligence as a single ability common to all intellectual processes.

(2) Those that regard it as a group of two or three abilities of varying degrees of generality.

(3) Those that regard it as representing no real entity but as merely a convenient term for average of all specific abilities.

Stern, Burt, Woodrow and Spearman are the followers of the first view.

According to Stern 'Intelligence is the general adaptability to new problems and conditions of life'.

Burt defines it as 'Inborn, all round mental efficiency'.

Woodrow calls it, 'a capacity to acquire capacity'.

Perhaps the most astounding definition and one that has given rise to a good deal of discussion and educational research is that of Professor Spearman. It was in 1904 that Spearman published his correlations between sensory tests and estimates of intelligence which showed that 'all branches of intellectual activity have in common one fundamental function (or group of functions) whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all others.' The one he called the general factor, symbolically represented by '*g*' and the others were termed briefly as S_1, S_2, S_3, etc. It is this general factor or 'common central factor' that measures native intelligence. Thus when we respond to any situation or perform an action, our general mental efficiency is reponsible for part of this reaction and our specific ability in that particular subject is responsible for the rest. Our performance in a physics paper, for instance, is due partly to our general intelligence and partly to some kind of special ability in physics which we somehow possess. This theory of Spearman is called the 'two factor theory of Human Ability'. Spearman could very well show how the amount of '*g*' involved in different test situations may be ascertained according to the set of correlation coefficient we get, or in more technical language, according to the hierarchy of correlations. The amount of '*g*' needed to meet different situations is not the same. Some require more '*g*' and others less. Hence some test situations are more saturated with '*g*' than others but according, to this analysis of Spearman there can be no situation in which some amount of '*g*' however small will not be needed. Spearman's two factor theory satisfactorily explained the tendency for all abilities to overlap to some extent, and yet to show considerable uneveness. The pupil who is good at English is usually superior also in arithmetic, and even at handwriting. At the same time each subject involves its own specifics, or 'S' factor. Hence some pupils may be relatively better at English than at number work, or vice versa. The specific factors in practical and physical abilities are larger and their '*g*' saturations are smaller. Hence discrepancies in these abilities are much more common. There is a good deal of truth in Spearman's contention though in practice it is very difficult to separate '*g*' and the 'S'.

Spearman examined the measurements of any four cognitive abilities—abilities that are concerned with thinking or intellectual aspect of the mind by calculating their coefficients of correlation. He arranged these coefficients in the following way and got some new conclusions.

$$(r_{ab} \times r_{cd}) - (r_{ad} \times r_{bc}) = 0$$

Where r_{ab} denotes the coefficient of correlation between the abilities *a* and *b*. Similarly r_{cd}, r_{ad} and r_{bc} are other coefficients of correlation between different abilities. He called this formula as 'tetrad equation'. The value of the left side is called the 'tetrad difference'.

The following illustration from Rex Knights' book, 'Intelligence and intelligence testing', will make the two factor theory clear.

Suppose that we have measured five cognitive abilities, the ability to draw inferences, the ability to complete sentences, the ability to continue series of numbers, the ability to supply synonyms of given words, and the ability to code messages. Let us suppose, further that we have calculated the coefficient of correlation between every one of the five abilities and each of the other four, and that these coefficient are as follows :—

	Inferences	Sentences	Numbers	Synonyms	Codes
Inferences	—	·42	·35	·28	·21
Sentences	·42	—	·30	·24	·18
Number	·35	·30	—	·20	·15
Synonyms	·28	·24	·20	—	·12
Codes	·21	·18	·15	·12	—

The table of coefficients satisfies the tetrad equation. If we take any four of the abilities, and name them as '*a*', '*b*', '*c*', and '*d*', we find that $(r_{ab}\times r_{cd})-(r_{ad}\times r_{bc})=0$. For example let '*a*' denote the ability to code messages, '*b*' to ability to complete sentences, '*c*' the ability to draw inferences and '*d*' the ability to continue series of numbers. Then the tetrad equation becomes $(\cdot 18\times\cdot 35)-(\cdot 15\times\cdot 42)=0$.

Similarly let '*a*' denote the ability to supply and synonyms '*b*' the ability to continue series of numbers, '*c*' the ability to draw inferences, and '*d*' the ability to complete sentences. Then the tetrad equation becomes $(\cdot 20\times\cdot 42)-(\cdot 25\times\cdot 35)=0$.

We have considered only two sets of four abilities, but any set of four will be found to satisfy tetrad equation. By mathematical reasoning Spearman claims to show that since measurement of different abilities satisfy the tetrad equation there must be in each of these abilities a general factor, which is common to them all, and a specific factor, which is confined to the particular ability in question. This is the two factor theory. A little reflection will show that abilities which are each divisible into two factors, general and specific must satisfy the tetrad equation. Every such ability since it partly depends on the general factor, is positively correlated with it to a greater or less extent ; abilities that depend much on the general factor are highly correlated with it, while those that depend most on their specific factors are only slightly correlated with the general factor. Further, since any two such abilities are both correlated with the general factor, they must be correlated with each other. If for example, the ability to draw inferences is correlated with the general factor to the extent of ·7, and the ability to complete sentences is correlated with general factor to the extent of ·6 then the ability to draw inferences is correlated with the ability to complete sentences to the extent of ·7×·6=·42. Suppose now we have five abilities—the same five as we have considered before, the ability to draw inferences, the ability to complete sentences, the ability to continue series of numbers, the ability to supply synonyms, and the ability to code messages. Suppose further that they are correlated with the general factor to the extent of ·7, ·6, ·5, ·4 and ·3 respectively. Then for example, the ability to complete sentences will be correlated with the ability to supply synonyms to the extent of ·6×·4=·24 and similarly others. In short the coefficients of correlation of the various possible pairs will be those set out in the table on the previous page which, as we saw, satisfies the tetrad equation.

Spearman's proof of the two factor theory moves backward from this result. He claims to have shown not only that abilities which are divisible into two factors, general and specific, necessarily satisfy the tetrad equation, but also, conversely, that abilities which satisfy the tetrad equation are necessarily divisible into two said factors.

For us, the important feature of the two factor theory is its statement that a general factor enters all our cognitive abilities and underlines all our thinking. This general factor is important because we can identify it with intelligence. Spearman himself names it '*g*'. He thinks that the word 'intelligence' has become 'a mere vocal sound, a word with so many meanings that finally it has none, and he tries to avoid its use. Still he does explicitly identify '*g*' with intelligence in more than one passage, and, as we shall see, this identification is fully justified.

Spearman claims that the intellectual activities depend on general and specific abilities. But he agrees that the specific factor may contain elements that are present in some others, he admits that, in certain instances, specific factors to some extent overlap to form broader functions or 'group' factors. Thus the specific factor, the special ability, which in addition to general intelligence, is required in learning French partly consists of verbal ability which is a group factor that enters into all mental operations involving words. The most firmly established group factors appear to be, (1) Verbal ability, (2) Numerical ability, (3) Mechanical ability, (4) Musical ability, (5) Logical ability, (6) The capacity for sustained intellectual effort, (7) Certain forms of memory, (8) The ability to change quickly and effectively from one mental task or train of thought to another, (9) Group factors in types of school work as literary, scientific and manual.

The second theory, *i.e.*, the group theory is held by Binet, who proposed that intelligence is made up of the following abilities :

(1) The ability to comprehend a problem, directing the mind towards its solution and maintain attention upon it.

(2) The ability to adapt the mind to the needs of the situation.

(3) The ability for self-criticism.

Maxwell Garnett also holds the same view of group nature of intelligence.

He says "It is probable that genius, as the word is commonly understood, is more directly measured by $E=\sqrt{g^2+c^2}$ than by 'g' (measuring general ability or capacity to concentrate attention) alone, or even by c (measuring cleverness as we define it, or tendency to associate by similarity). He concluded that the general factor 'g' is compounded with one or more group factors to constitute intelligence.

The third theory, the theory of independence and several traits that go to make up the mind, has been advocated by Thorndike, Woodworth and G. H. Thomson. According to Thorndike intellectual ability consists of the power to form numerous 'bonds' and each situation requires a bundle of these bonds or elementary capacities. He denied that there is any 'common central factor'. Woodworth and Thorndike proved in a series of brilliant experiments that the transfer of training (for example, the effect of learning of poem on the learning of another, or the learning of Latin upon the learning of other school subjects) is far less than educationists supposed it. This led to the statement of the view that learning is not general but specific and that mind is composed of a host of highly particularized and independent faculties.

Some of the Spearman's statistical techniques were strongly criticized by G. H. Thomson, and he argued that the two factor theory was a possible, but not a necessary, inference from the statistical results. Thomsons Sampling Theory assumes the mind to be made up of many independent powers. Any particular test or school examination samples some of these powers or bonds. In so far as two tests sample the same bonds then a general common factor can be said to exist between them, but in so far as they sample different bonds then the tests have nothing in common and each is specific.

After a long period of experimentation for investigating the human abilities, Thurstone gave the results of the analysis of 56 tests, given to 240 college students. Thurstone's analysis shows that the best way of interpreting the correlation data of different abilities is not merely to suppose a '*g*' ability and different 'S' abilities but to suppose a finite number of a factors or abilities of which '*g*' may be the first and most important. The eight main or primary factors were indentified as :

V = verbal comprehension
I = Inductive reasoning
M = Rote memory
W = Word fluency
P = Perceptual speed
N = Number
D = Deductive reasoning
S = Space or visualization

Certainly Thurstone did not disprove the existence of '*g*' but in his technique of factor analysis he could not find '*g*'.

According to him the human ability is best explained by a number of other abilities of minor importance. Those minor abilities may, however, be of considerable importance in certain walks of life where that particular type of ability may be more needed than the general intelligence '*g*'. For instance in the performance of a mechanics tasks the practical factor may be much more important than '*g*'.

The story of the theory of intelligence does not end here. There are many other psychologists like Ebbinghaus, Terman, Colvin, Piaget etc., who define intelligence in their own way. It will not be out of place here to mention the definitions of these psychologists.

Ebbinghaus says, 'Intellectual ability consists in the elaboration of a whole into its worth and meaning by means of many sided combination, correction and completion of numerous kindred associations. It is an activity of combination.'

Terman says, 'An individual is intelligent in proportion as he is able to carry on abstract thinking'.

Colvin defines it in terms of behaviour and calls it 'a group of innate capacities by virtue of which the individual is capable of learning in a greater or less degree in terms of these innate capacities with which he is endowed'.

A novel view of intelligence is held by Piaget. He writes that 'Behaviour becomes more intelligent as the pathway between the subject and the object on which it acts ceases to be simple and becomes progressively more complex. Less obvious relationships are only perceived by the more intelligent person. Only the unintelligent are naive'.

Thus we have seen the divergence of opinion as regards the definition of intelligence.

All these definitions can be summarized in a few lines as below :

"Intelligence is a general mental ability which operates in many different ways, is more fully manifested in the higher mental processes than in the lower ; it is specially active in dealing with a situation which presents points of novelty (in other words with the solution of problems'. It is more concerned with the dissecting, planning and rearranging of the data of experience than with mere reception of impression".

The Measurement of Intelligence

It is a matter of day-to-day experience that some pupil learn one thing more easily and accurately which a fewothers cannot do inspite of the fact that they receive the same instructions. It is all the more surprising to note that some pupils with a better opportunities of instructions lag behind than to those who receive a lesser coach. This ability which makes the pupil to learn more easily, quickly and accurately is nothing but intelligence. The educational career of a school child depends upon his 'native intelligence', his 'physique', his 'deligence', 'the environments in which he is brought up', and lastly 'the length of his schooling'. The success in the school career depends upon all these factors. The teacher, the parent and the medical man can help to improve the abilities except the first one, *i.e.*, the native intelligence. In the case of a child born abnormal what the utmost we can do is to improve the situation slightly. There are certain difficulties in measuring intelligence. It is difficult to separate pure native intelligence from acquired knowledge.

As said before Alfred Binet was the first man who attempted to measure intelligence in the year 1905. The first intelligence scale was produced by Binet in collaboration with Simon in Paris. In 1908 the intelligence scale was issued in a revised form, and finally in 1911, the scale as we have it at present was published. Some tests which had been found unsuitable in the previous scale were dropped.

After a long period of research Binet discovered that the work of a teacher, who teaches normal children, becomes easy. They can learn easily and quickly. Till that time teaching techniques have progressed much. People knew which type of instructions should be given to the children of various ages. But it was still difficult to suggest the type of instructions in the case of mental defectives. A twelve year old frequently has to be taught like a eight year old, or like a seven year old or even like a five year old. The problem before Binet was as how to treat each child. It was by considering the mental defectives to be mentally like a much younger child that gave Binet his idea of measuring intelligence by comparing the activities and abilities of the children of one age with those of other ages. Galton, Wundt and Cattell were able, till that time, to measure the strength of memory, ability of recognizing various shapes and forms, span of attention etc. Binet could get some help from this work but he was well aware that intelligence is much more than the strength of memory or span of attention etc. His problem was to find out a normal child.

Binet discovered that by testing a large number of children (selecting them at rendom) and taking the average performance of the whole group, it is possible to find out the performance of a 'normal child'. Those children whose scores tally with the average score of the whole group of the same age would be the normal or average children.

Binet, together with his friend Simon, worked on this idea for many years until in 1908 they were able to publish a list of things one might except of normal child of various ages to do.

'The Test Items of Binet-Simon Test for various ages.'

Age 3

1. Pointing to nose, eyes and mouth.
2. Repetition of short sentences.
3. Repetition of two digits.
4. Enumeration of objects in pictures.
5. Knows his last name.

Age 4

1. Knows sex.
2. Names familiar objects.
3. Repetition of three digits.
4. Knows longer of the two lines.

Age 5

1. Compares two boxes of different weight.
2. Copies square.
3. Games of patience, Rectangular card cut diagonally to be reconstructed according to a similar uncut card placed before the child.
4. Counts four coins.
5. Repeats a sentence of ten syllables.

Age 6

1. Knows right and left.
2. Repetition of a sentence of 16 syllables.
3. Aesthetic comparison.
4. Definition of familiar objects.
5. Executes three commissions.
6. Knows age.
7. Distinction between morning and afternoon.

Age 7

1. Unfinished pictures, tells what is missing.
2. Number of fingers.
3. Copies a written model.
4. Copies a diamond.
5. Description of pictures.

6. Counts 13 coins.
7. Repetition of five digits.
8. Knows names of four common coins.

Age 8

1. Reads a passage and remembers two items.
2. Counts nine coins.
3. Names four colours.
4. Counts backward from 20 to 0.
5. Writes from dictation.
6. Differences.

Age 9

1. Knows date, day of week and month and name of month and year.
2. Repeats days of week.
3. Makes change (with coins).
4. Definition superior to use.
5. Reads a passage and remembers six items.
6. Arranges five boxes in order of weight.

Age 10

1. Repeats the month of year.
2. Knows the names of nine pieces of money.
3. Use three words in one sentence.
4. Comprehension of common sense questions.
5. Comprehension of common sense questions (difficult).

Age 11

1. Criticises absurd statement.
2. Uses three words in one sentence.
3. Names sixty words in three minutes.
4. Defines abstract words.
5. Disarranged words arrange into a sentence.

Age 12

1. Repetition of seven digits.
2. Find three rhymes for a given word.
3. Repetition of a sentence of 26 syllables.
4. Interprets pictures.

Age 13

1. Paper cutting in different shapes.
2. Reversed triangle, rearrange two triangles in imagination and draw results.
3. Differences between pairs of abstract term.

Binets method was to give these tests to each child individually. In quiet and congenial surroundings, the examiner set these questions to the child one after another with the help of necessary material aid. Binet tests of intelligence were, therefore, individual tests and were mostly oral. It has been found that this is the best procedure in mental testing because it enables the examiner to establish proper contact with the child and thus get a correct picture of child's reaction. The significance of this scale of Binet is that a really usable way of comparing the intelligence of various peoples was given to us. What one child can do in comparison with other children of the same age, gives the clue. It became clear that a ten year old child doing the tests of an eight year old child was retarded two years out of ten. A ten year old able to do the tests expected of the normal twelve year old child was obviously of the same mental ability or mental age as twelve year old boy. This is really a significant idea which Binet gave to us, about the mental age. We know a great deal about a child when we hear that the chronological age (his age calculated in years) is ten and his mental age is also ten. We know immediately such a child is about normal in mental ability. If the mental age is 10 and C. A.=8, we know that the child is above normal. If MA<CA, we know that he is retarded and if the exact ages are given, we know just how much he is retarded. Binet gave us, in the concept 'mental age', a way of classifying people according to their native ability or speed of learning. He could now classify mental defectives and plan their education more carefully.

The Binet Simon tests were adopted by different countries. In England in 1922 Prof. Burt issued the London Revision of the tests known as *mental and scholastic tests.* In America Terman carried out his revision between 1913-16 and named it *Stanford Revision and Extension of Binet Simon Scale for measuring intelligence* with a guide book called *the measurement of intelligence.* In Germany, Bobertaga issued his revision in 1913 and in Italy the psychologist Saffiotte issued a similar revision for his country. In English speaking countries, Terman's revision proved most satisfactory and is now generally recognized as the standard intelligence test for use in psychological testing. Terman who was working in Stanford University (America) in collaboration with Merrill brought out in 1937—'The new Revised Stanford Binet test of intelligence.' The new revision is a carefully selected scale which gives very reliable results, and is now in universal use both in England and America. In India, Bureau of Psychology, Allahabad, U. P., has also adapted Binet test for Indian children.

The method of Binet was scientifically defective. A retardation of one year in the case of a child of six years is more serious than the same amount of retardation in the case of a child of twelve years. It was also experimentally found that by the time a normal child of six years mental age grew up to the mental age of twelve years, a child of the same chronological age and five years mental age grew up roughly to one of ten years mental age. This means that a gap of one year at the lower level became one of two years at the upper level. Hence it was found that it was not proper to compare children merely by their mental ages. A child of the chronological age of four years having a mental age of three years was almost on the border of mental deficiency while one of sixteen years of age having a mental age of fifteen years was quite normal. Hence an absolute measure to compare children of different ages was required. It was German psychologist 'Stern' who gave the idea of 'intelligence quotient.' By some it is called Mental Ratio.

Intelligence quotient, symbolically called I.Q., is the ratio of mental age to chronological age. This ratio is multiplied by 100 to obtain an integral value of the quotient.

Thus $\text{I.Q.} = \frac{\text{M.A.}}{\text{C.A.}} \times 100$

Thus if M.A. of a child $= 10$

and his C.A. $= 10$

His I.Q. will be $\frac{10}{10} \times 100 = 100.$

An I. Q. of 100 denotes a normal intelligence and a person possessing an I.Q. of 100 is called to be a normal person. If M.A. is less than C.A. the I.Q. will be less than 100, *i.e.*, below normal. Again if M.A. is greater than C.A. the I.Q. will be more than 100, *i.e.*, above normal. Thus we have seen that in the case of a normal child M.A. = C.A.

The following table shows the classification of English pupils according to I.Q.

Standardised score to an I.Q. basis	Percentage of children within this group	Description	Approximately suitable secondary and higher education
145 and over	·2	Exceptionally Bright	Universities and higher education.
130–144 115–129	1·8 10	Very bright Bright	Grammar schools, technical schools and technical colleges, universities
110–114 85–99	38 38	Average	Secondary modern schools
70–84	10	[2] Dull and backward	Special class in secondary modern school
55–69	1·8	Mentally deficient	Schools for educationally subnormal (E.S.N.)
Below 55	·2	Idiots	Institutions (ineducable in the ordinary sense)

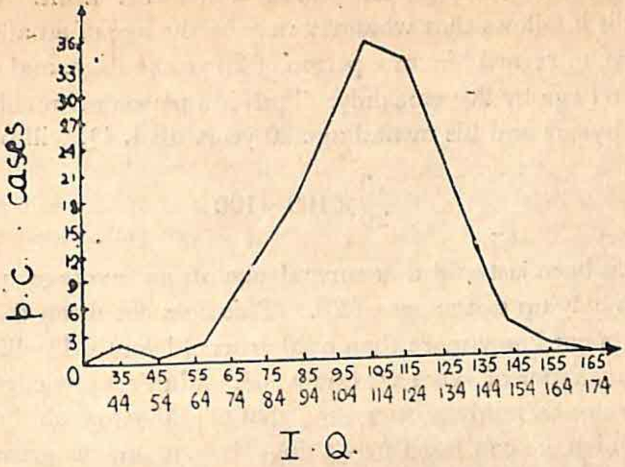
The classification of the children of our country is as follows:

Class	Range of I.Q.	Percentage of General Population	
		Actual	Smoothed
Near genius or genius	140 and above	2 00	1·5
Extraordinary	130–139·9	3·62	3·5
Very superior	120–129 9	7·10	9
Superior	110–119 9	14·43	15
Average or normal	90–109 9	42·73	42
Backward	80–89·9	15 50	15
Very backward	70–79·9	10 50	9
Border line	60–69 9	3·20	3·5
Mental Defectives:			
Morons	40–59·9	0·93 (morons only)	1·5
Imbeciles	20–39 9		
Idiots	Below 20		
	Total	100	100

---

[1] The pupils for grammar and technical schools are paired here, but the method of selection of pupils for these schools varies considerably.

[2] Sometimes the backward child is said to be one whose attainment quotient is below 80 (See Reading Ability, H.M S.O. pamphlet).



Graph showing the distribution of I.Qs.

The data of the distribution of I.Q. are presented graphically in the above figure. It will be noted that a close approximation to a normal curve is attained in developing these scales.

Individuals possessing intelligence between any two given levels can be represented by the area of the surface enclosed by the corresponding ordinates (vertical heights) and the boundary of the curve. We have thus a continuous transition from the dullest to the most intelligent in human group.

The growth of mental age

The I.Q. of a person is assumed to be same all throughout his life. It has been found that the I. Q. undergoes some changes during the course of the years of the individuals life but for practical purposes the layman may well assume the I.Q. to be comparatively stationary throughout life. Now although the I.Q. of a child or an individual is constant at different ages, the mental age is not. The mental age of a child goes on increasing, so that the same child has for example a different (higher) mental age at 10 years of actual age from what he had at 8 years. Now it has been found that this increase in mental age takes place upto 20 years of age. In case of mentally inferior children this growth stops a year or two earlier, and in case of mentally superior children, it may continue for a year or two at a very slow rate. However 20 years is taken as a limit. If we adopt this it follows that whatever may be the age of an adult we shall have to regard him as a person of 20 years of age, and divide his mental age by 20 years only. Thus if a person's chronological age is 25 years and his mental age 20 years his I. Q. will be

$$\frac{20}{20} \times 100 = 100$$

As has been just said that mental age of an average person increases only up to the age of 20. This does not mean that the adult does not know more than a 20 years old youth. Intelligence is not a matter of what we know. An adult's knowledge and maturity are certainly greater than that of 20 years old but the rate at which we can learn new things is not much greater at 25 or 40 years of age than at 20 years of age. Intelligence means ready adaptability to new situations, alertness, keenness, ingenuity. It is quite different from knowledge and experience. We by no means reach our maximum power at 20 (we reach these powers nearly at 50 if we avoid mental laziness) but the normal youth, if placed in a situation in which previous experience gives no advantage, has a chance to meet the situation just as intelligently as the normal adult.

Administration of Binet test

For administering the Binet test the psychologist first wins the confidence of the subject and makes him feel at ease. One way in which this may be accomplished with young children is to suggest them that the tester and the child are playing a game. The tester also encourages the child by praising his performance. Such encouragement is given whether the responses are right or wrong. The child is never told 'that is wrong'. He is encouraged by such comments as 'That's very good' now we will try the next one ! Before administering the actual test, the tester tries to form a preliminary estimate of the subject's intelligence. For instance the teacher's estimate of intelligence can be taken to give the rough idea about the intelligence of the child. Then he starts testing a little below the estimated level, and works up the scale till he reaches a particular level at which the subject is not able to pass a single test. Suppose that we are to test a child of 10 years and 1 month old reading in class IV, having the rank in the class 15 out of 25. Let teachers estimate of his intelligence, be 'D' out of 'A', 'B', 'C', 'D', 'E'. The calculation of the mental age is now very simple. Start with a age at which the child answers all the questions as the basic age. Then in the higher ages since there are six items in each year, credit two months for each item scored right. After the tenth year the age groups are XII, XIV, XVI, XIX and XXII. This means that in the range of two years from tenth to twelfth year there are only six items. Hence for each item scored right credit 4 months. The same is the case with the tests in the XIV and XVI year groups. In the next higher groups there are 6 items over a range of 3 years. Hence credit 6 months for each item scored right in the group XIX and XXII. Thus if the above child whose C.A. = 10 years and 1 month could answer all the items of year VII, 3 items of year VIII, 2 items of year IX and no item of year X. We may calculate his M.A. as follows :—

All items of year VII	Mental age 7 years.
3 items of year VIII	6 months.
2 items of year IX	4 months.
No items of year X	
Total	M.A. = 7 years 10 months.

∴ M.A. = 94 months.

∴ C.A. = 121 months.

$$\therefore \quad \text{I.Q.} = \frac{94}{121} \times 100 = \frac{9400}{121} = 77{\cdot}7$$

His C.A. = 10 years and 1 month and M. A. = 7 years and 10 months. The child is therefore retarded by 2 years and 3 months.

Types of Intelligence Tests

There are two broad classes of tests which are used today for measuring intelligence : individual tests and group tests. Individual tests follow in the tradition of the well-known test first devised by Binet and Simon to measure the intelligence of mentally defective children. This test was later developed for measuring the intelligence of all children and has been amended and revised several times since. Notable revisions are those of Burt and Terman and Merrill. The latter revision is very widely used in English-speaking countries. On the continent psychologists are preparing standardisations for German and Scandinavian populations. Valentine has also produced an individual test of intelligence which has much in common with the various revisions of the Binet test. Most performance tests, Kohs Blocks, Cube Construction, Alexander's Passalong, the form boards, and a Blocks test devised by Dr. Peel to mention only a few, are also individual tests.

(1) **Group Tests**—Group tests have a shorter history. The entry of the United States into the First World War is usually accepted as the beginning of group testing. The Americans were faced in 1917 with the problem of testing many recruits quickly, and devised the group test to meet the difficulty : that is, a test which can be taken by a large group of people at once. Very soon group tests 'caught on' in the testing and grading of school children. In Great Britain Thomson and his co-workers at Moray House have been the biggest agency for promoting the widespread use of group test material. Thomson and Burt's original Northumberland Tests were first issued in the early 1920s.

Since the group test of intelligence enables us to test large groups of children at once it is now almost universally used by local education authorities for purposes of grading and selection of pupils for different forms of education. For the same reason it has found wide use by the military and by research workers investigating the qualities of intelligence and predicting aptitudes.

The tests normally require 20-60 minutes to do, for most are timed. Very often the items of the test are arranged in order of difficulty. Broadly speaking there are two forms in which a group intelligence test can be set out : the omnibus form in which the testee proceeds through the test to a time limit set for the whole test ; or the sub-test form in which the test is really made up of several smaller tests, each of which is timed separately, and the items of which are separately arranged in order of difficulty. Each sub-test usually contains one kind of item, say, one made up of analogies, another of reasoning problems, and so on. There are advantages and disadvantages in both kinds of test. In the omnibus test the child is left in peace for the whole period of testing whereas in the sub-test form he has to be interrupted at the end of the time allocated to each sub-test in order to start him on the next. It is generally not considered desirable to have sub-tests of less than five minutes' duration. On the other hand evidence has been produced that the long period required for the omnibus form leads to fatigue and boredom which is reflected in the test score. In the sub-test form this element is not so evident and we can more safely assume that a child's score is free from fatigue effects and reflects his real intellectual limits.

Another way of classifying intelligence tests is the speed test or the power test. A power group test is one in which the items are arranged from the easiest to the most difficult and in which the testee is given as much time as he requires to answer as many items as he can. (Cf. Thorndike's concept of altitude of intelligence). A speed test is normally composed of items of a similar degree of difficulty, and the testee is limited to a set time for the test. This concept of intelligence has something in common with the Thorndike concept of breadth of intelligence, particularly if the items of similar difficulty are drawn from different fields—verbal, numerical, non-verbal, etc.

In actual practice most tests used for selecting pupils for secondary education combine speed and power, that is, the items are arranged in order of difficulty and a time limit is set on the test. We may ask whether the time limit set on the tests may do injustice to the slow accurate plodder. Evidence obtained from correlating speeded and unspeeded tests shows that at 11+ the abler child is also the speedier child. Furthermore the time limit imposed on most group tests is not so restricted as to make speed an important element to anyone except the very slowest, and these children are not the ablest. Also when viewed as a prognostic tool, the test should take speed into account, for the pupil in the secondary school or the apprentice in a vocation has to be able to work reasonably quickly.

There is evidence of a speed factor in simpler motor and mechanical activity and it also seems that this factor is more evident when older people are tested, but it is very difficult to get anything like a consistent description of the extent to which speed is a separate factor in the hierarchy of mental abilities. In Sweden the problem is also being investigated in connection with finding tests which do justice to the Swedish rural population in the north of the country. Opinion is that speed is a cultural factor linked with urban and industrial life and that the countryman, by the nature of his life and work, may tend to be slower.

The group test can be used with advantage for normal children who have a reading age of not less that eight years. For children whose reading attainment does not reach this level the printed verbal instructions of the group test may prove too difficult. But this difficulty might be overcome by presenting the material orally as is done in the Moray House Picture Test and in the orally presented test devised by Cornwell. It is a very curious fact that the rubrics of instructions in verbal tests may act as a barrier and that if these details are read aloud to the children they perform much better on the verbal material of the tests.

Another reason for not giving a group test until a child has had a year or so in the primary school is that in the group test we take for granted that the child is sufficiently regimented by classroom procedure to carry out written instructions. The advantages of the standardised objective group test are that ambiguous questions have been eliminated in the preparation of the material, that the children's answers can be marked objectively and quickly. Also by the uniform conditions of group testing the atmosphere is the same for all. The uniformity may, however, be illusory and even irrelevant, for the important thing is that each child should feel at ease so that his test score is a measure of intelligence and not a reflection of emotional tension.

Group tests of intelligence may be verbal, pictorial or symbolic and spatial. In the items of a verbal test use is made of words and numbers.

The chief means of social communication and the chief advance of knowledge is through the use of words. Hence it is most important to be able to assess verbal ability. But there are cases where the verbal test would be of less use. Such occur when we suspect that a child's development and schooling have not been normal. It will be obvious to the reader that the verbal test of intelligence makes use of learned material, words and numbers. The underlying assumption in the use of verbal tests is that every child tested has had a normal background and schooling. If this condition is not fulfilled then the test is invalid and may give rise to a false measure of intelligence.

The reliability of verbal tests is generally very high. ✓

Group pictorial tests may be used for young children and non-readers. They involve verbalised thought, through the medium of pictures, but do not require reading attainment. They are particularly useful for testing children in the age range 7—10. ✓

Group non-verbal tests make use of abstract geometrical patterns and designs.

Some non-verbal tests merely claim to measure general intelligence '*g*', others general and spatial intelligence '*g—k*'. The amount of '*g*' measured by a non-verbal test is rarely so high as that measured by a verbal test.

Non-verbal tests may be used whenever we wish to assess intelligence apart from the verbal influences and whenever we have reason to suspect that a person's background and schooling have not been normal, or if a child suffers from any reading disability. They are also better to use when reading attainment is to be compared with intellectual capacity. They are preferable to pictorial tests for older pupils and adults since they may be made as difficult as we please.

(2) **Individual Tests**—Group tests are objective in marking but are not sensitive to the childrens' emotional differences. These differences may effect the test scores. Individual tests are slower to use but give the tester a much clearer impression of

the pupil being tested, and his emotional responses may be observed very closely and responded to accordingly. The individual test is therefore to be recommended whenever results of group testing are in doubt, for borderline cases or abnormal cases. The individual test is also to be preferred for younger children who cannot read or who are not accustomed to the regimentation and group methods of the class-room. The tester may thus eliminate elements which are irrelevant to the purpose of intelligence testing.

The Terman-Merrill revision of the Binet test is very frequently used for testing general intelligence and is issued in two parallel forms, L and M, which contain memory items, memory for numbers and sentences, reasoning items, vocabulary tests, following instructions, and so on.

Intelligence Tests Abroad

The following is a brief description of some of the intelligence tests used in the foreign countries. For the sake of convenience they have been classified under different school levels.

Tests for the Primary Level

Tests for the primary level require no reading or writing on the part of the subject. For this reason, they are sometimes described as non-verbal tests. This category should not be confused with the non-language tests. The latter type of test requires no language at all, either written or spoken, and is suitable for foreign-speaking and deaf as well as for illiterate subjects. The usual primary group test involves extensive use of spoken language one of the best known group tests for the primary level is the Pintner-Cunningham Primary test, which has been in use for approximately twenty years. Another test designed for the primary level is the Otis-Alpha. The test is available in two equivalent forms, A and B. Each includes a 'non-verbal' and a 'verbal' test. By an ingeniously devised plan, the same items are re-used for both verbal and non-verbal parts. Each part may be employed and evaluated separately, or the two may be combined to yield a single total score.

For the non-verbal part, the subject is instructed to mark in each set of pictures the one which does not belong with the others. A common defect in many such items is the possible ambiguity of correct response. It is difficult to prepare items in which only a single answer is defensible.

Tests for the elementary school level

These tests are designed for use from grade 4 through grade 8 or 9. Since literacy is presupposed at these levels, such tests are predominantly verbal in content ; most also include arithmetic problems or other numerical tests. One of the first group tests constructed for use with school children was the National intelligence test. This test was prepared shortly after the terminal of World War I by a group of psychologists working under the auspices of the National Research Council. Otis Beta was also designed for grades 4 to 9. The content of this test is largly verbal. The Otis Beta is a revision of the earlier Otis Self-Administering Tests of Mental Ability. All instructions are printed on the test booklets together with sample items of each type. Timing is also simplified through the use of a single time limit for the entire test.

Tests for high school students and unselected adults

Among the tests available for high school students, mention should first be made of the upper levels of the Pintner, Otis. The Pintner advanced test is designed for grade 9 and above. The Otis Gamma represents the third and highest level of the Otis Quick-Scoring Mental Ability Tests. This test follows closely the Pattern of the Otis Beta, substituting more difficult items of the same type.

Tests for college students and superior adults

Among the most widely used psychological tests for the college level are Scholastic-Aptitude Test prepared by the College Entrance Examination Board and the American Council on Education Psychological Examination for College Freshmen, generally known as A C E. The former consisting of a verbal section and a mathematical section, is restricted for exclusive use in annual testing programme conducted by the College Entrance Examination Board.

Each form of the A C E, on the other hand is released for general use after it has been employed in the annual testing of entering college freshmen.

Verbal group tests in England

Group tests of verbal form came in prominence in England after 1920. Prof. Godfrey. M. Thomson produced a number of standardized intelligence tests, the first under the name of Moray house tests. Burt also produced a few tests named Northumberland test, as these tests were first used in the country of Northumberland. Spearman produced those test items which have a high '*g*' saturation. Ballard devised tests known as 'Chelsea Tests' which have been extensively used in England. Recently Dr. Cattle has constructed some useful group tests. The National Institute of Industrial Psychology (London) produced a large number of useful group tests (both verbal and non-verbal) under the guidance of Charles Mayers.

Non-language Group tests and Performance Tests of intelligence

Owing to the deficiency in linguistic development the deaf are handicapped on verbal tests, even when the verbal content is visually presented. Similarly the child who has a serious speech defect or those whose speech development is retarded for any reason will be unable to take many of the tests on such a scale as the Stanford-Binet, which requires oral replies. Such a child, moreover, may be too young or too retarded educationally to take a written test. The illiterates are also unable to take such test. Children with special reading disabilities would also fall into the same category. Persons with a foreign-language background may like wise experience special difficulties in taking the usual predominantly verbal intelligence tests.

The performance and non-language tests have been developed to overcome these difficulties. Another purpose for which performance and non-language tests are employed is to supplement the usual type of intelligence test. Certain individuals score poorly on verbal tests for special reasons. Thus the shy, inarticulate child, or the child who feels discouraged, when confronted with verbal tests because of his school failures, may perform more satisfactorily on less academic tasks. On the other hand, the 'verbalist type' of individual may obtain a deceptively high score in certain verbal tests, although his understanding of most problems may be superficial and his practical judgment may be seriously deficient. It is now generally recognized that performance or non-language tests are not simply a substitute for verbal tests.

Non-language group tests

The first non-language group test was the Army Beta test developed for testing foreign speaking and illiterate soldiers in the United States Army during World War I. The Beta was given to all men who fell below a certain score on the Alpha. In this group were included not only those who were handicapped by foreign-language background or illiteracy, but also those who performed poorly on Alpha for any other reason.

Instructions for the Army Beta tests were administered by means of gesture, pantomime, and demonstrations on specially prepared black board charts. The examiners task was much difficult in Beta than in Alpha. In the construction of Beta, an effort was made to pattern it as closely as possible after the Alpha, since it was designed as a substitute for Alpha.

Several civilian revisions of the Army Beta were developed. A recent form is the 1946 restandardization of the Revised Beta Examination published by the Psychological Corporation. This form consists of six tests, including (1) Mazes, (2) Symbol-digit substitution, (3) Locating what is wrong in the pictures of common objects or of persons engaged in various activities, (4) Spatial relations, (5) Pictorial completion, and (6) Checking similarities between pairs of objects or numbers. Another well known non-language scale, designed for elementary school children, is Pintner Non-language test. This test was originally constructed for use with deaf children. Another example of this type of test is the Chicago Non-verbal-Examination. Although not as well known as either the Army Beta or the Pintner-language, this test is of interest because of the variety of subjects which it provides.

Performance tests of intelligence

In these tests the abilities to read or write in any language is not required on the part of the examinee. These tests usually involve the construction of certain patterns with the help of given materials or solving of certain problems set in terms of concrete materials. One of the earliest performance tests was the Form Board developed by Seguin for use with the feebleminded. Originally devised in connection with Seguins programme for the sensory and motor training of the mentally deficient, this Form Board was subsequently incorporated into a number of performance scales. In case of Form Board test a number of wooden blocks of different shapes are provided and the child is required to fit them into holes of similar shape in a board. Scoring is done on the basis of the time taken by the child and the procedure followed. An attempt to fit a block into a hole of different shape is regarded as an error.

Another performance test was developed by Porteus in 1924 and is called the Porteus Maze Test. The mazes can be administered with no verbal instructions by using the easier mazes for demonstration purposes. They range from the three-year to the adult level. The standard procedure is to have the subject trace with a pencil the shortest path from the entrance to the exit of the maze, without ever lifting the pencil from the paper. There is no time limit ; the subjects are not hurried in any way. If the error is made by crossing the line or entering a wrong path way the subject is stopped and given a second trial on an identical maze. If error is made on the second trial, a failure is recorded for that level.

Another performance test is 'Man Drawing Test.' In this case the subject is required to make a best drawing of a man he can with pencil on a paper. The scoring is done in terms of the important parts of the man, e.g., eyes, nose, hand, fingers etc., included in the drawing and in terms of the relations in which they are shown to stand to one another. The artistic qualities of the drawing are not taken into account. Age norms based upon the average performance in terms of the number of points out of a possible 50 or 80, at each age level, are available for deriving M. A. and I. Q.

Alexanders group of performance test—Dr. Alexander suggests a combination of three performance tests known as 'Kohs Block Design Test'; the 'Cube Construction Test'; and the 'Pass Along Test' as a battery for measuring intelligence.

(*a*) *Kohs Block Design Test*—It is a well known performance test of intelligence. It is essentially a test of mental analysis and synthesis and gives a fairly accurate idea of the individual's '*g*' and some indications of K (spatial). The test material consists of sixteen coloured inch-cubes and ten designs (in original Kohs Block, there are sixteen) painted in the same colours on cardboard oblongs. The subject is required to put the cubes together in such a way that the design may be reproduced in a specified time. The maximum possible score on this test is 25.

(*b*) *The Cube Construction Test*—It comprises of three parts—

(*i*) A 3″×1″×1″ model block with its four sides painted red but its top and bottom left unpainted, along with nine one-inch cubes so painted that the model can be reproduced by means of them.

(*ii*) A similar model block, this time with the top and four sides painted, but with the bottom left unpainted, along with nine cubes so painted that they can be assembled to reproduce this model.

(*iii*) A two-inch cube, unpainted along with eight one-inch cubes each with three of its four sides painted red. In each of the sets, the examinee has to construct a larger model block with the help of inch-cubes provided. Marks are awarded on the basis of the time taken to do so.

V.V.I (*c*) *Alexanders pass along test*—This is another well known performance test of intelligence. Here the subject is required to reproduce, in a specified time, a given design by passing along, not lifting the wooden pieces, coloured red or blue, which are presented in a standardized order, in a wooden frame. The whole test consists of a series of eight designs graded in order of difficulty. The maximum possible score on this test is 20. It gives fairly accurate measure of what Alexander has called 'F' or Practical ability.

Intelligence tests in India

It has been found that on an average, the people of some races and nationalities do less well in certain intelligence tests than the people of other races and nationalities. This had led some people to claim that the people of their own race or nationality are on the average more intelligent than the people of certain other races. Any such interpretation of the results of Intelligence test is based on the number of false assumptions. There is no intelligence test in existence that can show Indian children on the whole to be either superior or inferior in mental ability to the Chinese, French, English etc. Intelligence tests, now available can compare Indian children and American children only if the environment, education and culture of the two countries are nearly alike, which evidently are not.

In England it may be assumed that any child of nine years of age will have learned to tell the time even though this has not been taught in school. He sees so many clocks and everyone lives so much 'by the clock' that he finds it necessary to be able to tell the time and insist on being taught. In an Indian village there may not even be a clock (even some schools are without clock) and village life pays little attention to being 'on time'. Thus lack of ability in a child to tell the time does not reveal the lack of intelligence in India in the way it does in England.

Binet insisted that his scale could be used only on people living in a fairly homogeneous environments, and that as soon as it was applied to those of a different country, the standard or norm could no longer be considered valid.

It thus follows that any test for India must be adapted to Indian subjects and only Indian standards considered valid.

The first attempt in constructing the intelligence test in India was made by Dr. C. H. Rice of Lahore, who about the year 1922 published his '*Hindustani Binet performance point scale.*' The test were adaptations of the Binet test along with some additional performance test. The system of marking was such that a score in the scale was directly convertible into mental age. The norms of the tests were prepared on the basis of tests to 1070 boys of the Punjab between the ages of 5 to 16. The tests were primarily meant for the Punjab but they could be used in any Hindustani speaking area. But these tests needed many modifications.

The intelligence quotients obtained by using the Hindustani Binet Scale are by no means comparable to those Terman reports on the Stanford-Binet. One need only examine carefully the following table from Rice's books to see that the range of I. Qs. on the Hindustani is much greater. Therefore, the term I. Q. must be used very carefully in India. An I. Q. of 130 on an Indian Scale is not nearly so hopeful as the same I. Q. on an American scale nor is an I. Q. of 65 nearly so hopeless as the same I. Q. in some other countries.

% of cases		Classification	Standard I.Q.	Hindustani I.Q.
Highest	5%	Genius	140 and up	165 and up
Next	5%	Very Superior	120–140	140–165
Next	15%	Superior	110–120	120–140
Middle	60%	Average	90–110	85–120
Next	15%	Dull	60–90	70–85
Next	5%	Borderline	70–80	55–70
Lowest	1%	Feeble-minded	Below 70	Below 55

After this attempts were made at standardizing verbal group tests, some in U. P., some in Calcutta and others in Madras presidency, but none of these have had wide appreciation.

Dr. Kamat of Bombay issued his 'Bombay Karnatak Revision' of the Binet test in 1940, together with a guide book, '*Measuring*

Intelligence of Indian Children'. The work done by Dr. Kamat was thorough and on scientific lines and the revision issued by him is an eminently suitable one. These tests could be used by only Marathi and Kanarese speaking children.

In 1941 Major Sohan Lal of the Govt. Training College, Allahabad, took up the Standardization of group verbal tests of intelligence suitable for children between the ages of 11 and 12. It proved quite useful.

Of course before the tests of Major Sohan Lal a few other verbal group tests were constructed in India by psychologists. In 1927 Dr. J. Maury of Ewing Christian College published the verbal group test adapted to Indian conditions. It was named 'Preliminary Classification Test'. It was published in Urdu, Hindi and English.

In 1933 Pandit Lajja Shanker Jha published in the 'Teachers Training College Annual' of Benaras Hindu University a report of the examination of over 1000 pupils on a Hindu adaptation of the 'Simplex Mental Test' by C. A. Richardson. It is intended for children between the ages 10 and 18. Pandit Jha complains of linguistic difficulties on the part of the younger examinees. He thinks that the introduction of English into the school from standard first retards the linguistic development of the mother-tongue.

Pandit Jha also adapted Terman's 'Group Test of Mental Ability.' This test is intended for the same ages as Simplex and seems, from a superficial examination, to contain more appropriate material for the younger children of this age-range than the Simplex, although here again greater ability in reading than is warranted must be assumed.

S. Jalota of D. A. V. College, Lahore, has prepared a group verbal test for use among college students. It is prepared in Hindi, Urdu and English.

The Draw-a-man test was standardized for India by Emil W. Menzel. Norms are given for both age and school placement based on 2600 examinations.

In 1942 the "Union Christian Training College Non-verbal Group Test" (Berhampore, Bengal), was got ready of standardization. It was prepared by T. C. Vicary and M. Draper. It is intended for ages between 8—15. It eliminated the language difficulty and can be adapted for use in any language area.

It would thus appear that intelligence testing has not yet made even a proper beginning in our vast country. We need to have satisfactory adaptations of Binet's (individual verbal) tests, at least one for every major linguistic division of the country, group verbal tests of intelligence for the different linguistic divisions and for different ages of children and, finally, performance tests which do not require language ability, and therefore might be found particularly suitable for our country.

The Battery of performance test of Intelligence prepared by Dr. C. M. Bhatia is widely used in India. This battery consists of the following five tests :

(1) Kohs Blocks Design Test.
(2) Alexander's Pass Along Test.
(3) Pattern Drawing Test
(4) Immediate Memory Test for Digits.
(5) Picture Construction Test.

A detailed description of these tests has been given in Dr. Bhatia's book—'Performance Test of Intelligence Under Indian Conditions, Oxford University Press, Bombay 1955.'

Wechsler Tests

The work of Wechsler has been outstanding in the field of intelligence and hence a separate mention need to be made of his work here. In 1939 Wechsler published the results of his investigation on the measurement of adult intelligence. He pointed out the difficulty in the measurement of adult intelligence, especially in view of the fact that a number of the tests which are used for adults have not been well standardized.

Wechsler standardized his tests on 670 subjects from seven years to sixteen years of age, and on 1081 from seventeen to seventy years of age. In addition he examined a large number of cases for his sampling technique. After a great deal of preliminary work the following tests were chosen for standardization :

(1) An information test

(2) A general comprehension test

(3) A combined memory span test for digits forward and backward

(4) A similarities test

(5) An arithmetical reasoning test

(6) A picture arrangement test

(7) A picture completion test

(8) A block design test

(9) An object assembly test

(10) A digit symbol test.

Alternate—A vocabulary test

These were then combined to form the following four separate but interrelated intelligence scales :—

(1) The main individual Adult Examinations for 16 to 60, consisting of the first 10 tests listed but permitting a reduction of the number to as few as 7 depending upon their suitability to the subject.

(2) An adolescent scale for ages 10 to 16 consisting of the same tests but separately standardized.

(3) A performance scale consisting of five tests.

(4) A verbal scale consisting of five or six tests.

The material in Wechsler's tests is interesting and the score in terms of points allows for accurate comparisons of individual's abilities.

Uses of Intelligence Tests

Educational use. Intelligence tests are of great use to the teachers. Teacher's task is to handle the young growing minds.

He must, therefore, know as much as possible of these minds. Unless he possesses some accurate instrument he cannot diagnose and treat these minds. Such an instrument for the teacher is Binet's rod of mental measurement. Pupils vary in their ability to learn. If the same type of instructions are given to dull and bright boys, it will be only a waste of time. Children of very low mental calibre, for example those with an I. Q. below 60, might be separated out and special arrangements made for their education in special schools. Specially advanced or gifted children, those with an I. Q. above 140, might in the same way be separated and specially trained as they are most valuable assets of a nation. Even the children, between the two limits of I. Q., those between 60 and 140, could be divided into two or three parallel streams and instructions may be given to them accordingly This is possible only in large schools where different curricula could be arranged.

Now one may question 'why not the teacher can estimate the intelligence of a child, why he should do so with the help of mental tests. Certainly teachers estimate of intelligence is not accurate. In estimating the intelligence of a child the teacher does not take into account his age. Suppose two boys of different ages, i.e., 6 years and 8 years are tested in a class. If their scores in that test be equal we cannot say that both possess the same level of intelligence. Considering the age factor we can say that the former is more intelligent than the latter. Another thing is that teachers predict the intelligence of a child by his progress in school subjects. This is not a correct method. One child may be comparatively dull but more industrious and thus may score more marks than another child, who may really be more intelligent but less industrious. Lastly, in assessing the childrens' progress in their school subjects by their teachers, a good deal of subjective element creeps in, so that a child in whose favour a teacher happens to be prejudiced secures more marks than another, who is not so favoured.

Now the question is whether children should be classified according to their I.Q. or according to their M.A., for being separated into different categories for special instructions. Theoretically it appears to be better to classify them according to their mental ages. For a child may have a high I.Q., say 125, yet his mental age may be too low for the standard in which he is and he may be doing poor work. Another child of a lower I.Q. say 100 may have a higher mental age and may be doing much better work in the same class. The following are the illustrations of these two types of children—

$$\text{First Child} \begin{cases} \text{M.A.} = 12{\cdot}5 \\ \text{C. A.} = 10 \end{cases}$$

$$\therefore \qquad \text{I.Q.} = \frac{12{\cdot}5}{10} \times 100 = 125$$

$$\text{Second Child} \begin{cases} \text{M.A.} = 13 \\ \text{C. A.} = 13 \end{cases}$$

$$\therefore \qquad \text{I.Q.} = \frac{13}{13} \times 100 = 100.$$

Thus it seems in classifying children it is the M.A. that is more valuable than the I.Q. There are many practical difficulties. We know that the advanced child and the backward child advance with different amount of mental age each year. A normal child may advance by one year of mental age each year ; a backward child may advance by less than a year of M.A., and an advanced child may advance by more than a year of M.A. Thus though they may start together at the beginning of their careers it will be seen that in a few years their progress will be quite different.

Besides this the progress of different children of the same mental age may not be the same in different subjects, for a child may develop a special aptitude or liking for one subject rather than for another and the result will be a vast difference in the progress he makes in the two subjects. Again the physical condition of children, and their home environments may play an important part in the efforts put forth by them. Therefore even if we start by putting children of nearly the same mental age in the same class we shall soon have to differentiate between them, for some children advance more rapidly than others. Therefore in classifying children we should not do so according to their mental ages.

If a child is above average in intelligence, that is if his I.Q. is above 100, it is expected that his E.Q. should be above 100. If it is less it means that the child is not working up to his natural level. If it is more it means the child is working beyond his powers. Coming to the Accomplishment Ratio, if the child's E. Q. is 80 and intelligence quotient is 100, his Accomplishment Ratio will be 80. This means that a child is not working up to his capacity. On the other hand, if the child's E. Q. is 120 and his I. Q. is 100, his Accomplishment Ratio will be 120, which means that the child is working much beyond his powers.

A quotient of 100 is considered the goal. For example, if a pupil whose M.A. is 10 years has an E.A. of 9 years 2 months, his A.Q. will be $\frac{110}{120} \times 100 = 92$. In like manners a second pupil might have the same E.A., 9 years 2 months, but an M.A. of only 8 years 3 months. His A.Q. would be $\frac{110}{99} \times 100 = 111$. The interpretation of the first case, 92, is that the pupil is not living fully up to his capacity, which seems reasonable enough, human nature being what it is.

But the interpretation of the second case 111 is rather absurd, since it appears to imply that this pupil has exceeded what he is capable of doing by 11 per cent. A far more probable explanation is that the quotient is due to inaccuracies in the tests, and that in this case the errors in the achievement score were in the direction of making it too high, where as the errors in the intelligence score were in the direction of making it too low. The resulting quotients has added these errors. If the errors had been in the same direction, they would have tended to off set each other.

Vocational Guidance. Intelligence tests are also used in vocational guidance. It is the duty of the teacher or the parent to guide the child in such a way which may help him to be successful in his profession. For each profession or trade the minimum I.Q. required for successful work is determined by the actual observation of successful workers. Thus persons possessing I.Q. higher than 125 are suitable for teachers profession. The skilled trades, and the clerical professions require men of medium intelligence. Unskilled labour may use men of the lowest I. Q., even men below I.Q. 80. The psychologists could record the different I.Q. required for different professions. Thus with the knowledge of the I.Q. of a certain child we can direct him to that line which he should take. The employer's task in selecting a candidate for a post where intelligence is desideratum, is analogous to that of school principal in selecting applicants for admission to his school. An intelligence test will indicate which candidate is likely to adapt himself best to new conditions of work and if he is qualified in other respects, this is the candidate whom the employer will choose.

As has been said before, according to Thurstone there are many factors, e.g., V, F, N etc. in our activities. If we can isolate them in the abilities of individuals and also in certain vocations, it is claimed that we can guide individuals for their vocations with a great deal of certainty. The lesser will there be the number of factors analyzed the greater will be their utility.

The stage at which vocational guidance should be given. Talking of our own country, we see a good many children drop out of school and students out of colleges at various stages of their career. Some leave the school at an elementary stage, some at a higher secondary stage and some in the middle of the university stage. It is therefore, felt that different types of vocational instructions should be introduced at different stages. Specialised instructions in the trades should be given according to the stage. For example after the completion of the high school course there should be technical courses in many of the highly skilled trades.

Facts about intelligence

Intelligence and Physique. Damage or disease of the brain has long been known to cause mental disturbance. In general paralysis, for example, where there is a gradual decay of the nervous system, the patients mental life is reduced. Improper functioning of the endocrine glands also effects our mental capacity. There is evidence that other physical defects impair one's intelligence. Dawson applied intelligence test to 1077 children in the Royal Hospital for sick children in Glasgow. He concluded that it was only in cases where there was disease of ductless glands or of the brain that there was appreciable departure from the normal in intelligence.

Intelligence and sex. Intelligence tests have failed to disclose any significant differences in intelligence between sexes. Burt did indeed find that among nearly 3,000 London children, the girls excelled the boys at almost every age from 3 to 14. The difference he says swells to a maximum about the age of seven. At ten it is reversed in favour of boys, but their recovery is transient, towards 14 the superiority of girls is again visible. And he thinks that it supports the view that in intelligence, as in other respects, girls mature more quickly than boys but on the whole, he does not regard the difference as significant.

Now considering the difference in intelligence between boys and girls of our own country we find a contrary thing. Our girls live more in the company of mothers most of which are illiterate. They are given the training to become good mother and housewives. Boys mostly live in the company of fathers who are more literate than mothers. Much emphasis is laid to bring up the boys for a literary career. The environments in which the boys and girls of India deal are very different from those of other countries and naturally sex differences of our country are varying from those of other countries.

Dr. V. V. Kamat's study could reveal the following facts :—

"It will be seen that the boys and girls start together, but as they grow older the differences between them become wider and wider. The girls are throughout progressive at a lower level until about the eleventh year, when their curve again shows a steeper ascent and overtakes that of boys at about the 13th year. They again fall back slowly at higher ages."

The following table indicates the Kamat's findings.

C. A.	Boys		Girls	
	Number	Mean M.A.	Number	Mean M.A.
2 to 2 11	20	2·7	15	2·6
3 to 3 11	20	3·58	20	3·46
4 to 4 11	25	4·49	16	4·47
5 to 5 11	39	5·57	24	5·04
6 to 6 11	28	6·38	37	6·22
7 to 7 11	44	7·39	50	6·72
8 to 8 11	35	9·16	46	8·16
9 to 9 11	39	9·79	41	8·94
10 to 10 11	61	11·3	42	9·15
11 to 11 11	55	11·55	38	10·65
12 to 12 11	54	12·67	31	12·08
13 to 13 11	50	14·0	32	14·06
14 to 14 11	38	14·53	18	13·73
15 to 15 11	44	15·86	13	14·83
16 and above	86	15·76	12	14·8

The influence of extraneous factors on scores in Intelligence Tests

Influence of practice. The effect of practice on the results of an intelligence test may be shown by giving two similar tests at an interval of, say, four weaks of each other. If this is done any thing up to an average rise of five points of I. Q. may be obtained

Influence of Coaching. Coaching in intelligence tests involves going through test instructions explaining the kind of items used, giving practice in the items and explaining the principles underlying the items. The effect of coaching are very marked, ranging anywhere from nine to eighteen points of I. Q. according to the brightness of the child and his lack of previous experience.

Effects of Schooling. In 'Sweden', Husen has shown that normal schooling may produce a rise of something like five to seven points of I.Q. According to Watts, Pidgeon and Yates, continued schooling in the secondary school produces a fall in I. Q. whereas if the child continues in a primary school there is a rise in the I. Q. Broken schooling and shortened schooling also has a marked effect on test results.

Cultural influences. It is difficult to construct a culture free test. The cultural influences which effect the test are language, way of administration of a test, traditions and manners of the country.

Intelligence and Heredity

Mental deficiency has been considered by some investigators as a recessive characteristic, that is, that the inheritence of less than normal mental capacity is due to germ plasm deficiency. Geneticists have shown that certain structural and disease entities are inheritable. Claims have also been made regarding the inheritance of such conditions as some types of insanity and susceptibility to diseases.

Recently studies of phenyl-pyruvic acid in the urine of mental defectives have created an increased interest in the genetic study of mental deficiency. In normal persons phenyl-pyruvic acid is never found. This study of Jervis let him to believe that there is transmission through a single recessive gene.

One of the best methods of studying the hereditary factors affecting intelligence is the investigation of the development of twins. We have sufficient information to know the probability of differences between fraternal and identical twins. Thus, for example, if we study twins reared apart under different environmental conditions, we may be able to make some definite conclusions regarding the effect of the different environments. In the same way significant data may also be obtained from the studies of foster children from the same family placed in different homes.

In 1905 Thorndike made the study of twins. He found on the examinations of 50 pairs of twins that the coefficient of correlation of their capacities was between ·6 and ·9, and of siblings about ·3. Older twins were no more alike than younger ones, and therefore it may be concluded that the environmental factor was not as significant as the hereditary factor. Sandiford could show from the following table the coefficient of correlation for intelligence among groups exhibiting degrees of blood relationship.

Group	Coefficient of Correlation
Identical twins	·90
Like sex twins	·82
Unlike sex twins	·59
Siblings	·50
Cousins	·27
Unrelated children	·00

From this table we can say intelligence is an inherited trait. Many say that similarity in intelligence is due to the similarity of environment. We can criticize this view for the following reasons :

1. Twins inspite of being placed in the same environment show greater degree resemblance.

2. Twins are alike in intelligence at the age of 5 as they are at the age of 16.

3. The differences in intelligence are found among children of different heredity who have been brought up in the same orphanage and attended the same school.

4. Evidence of influence of heredity on intelligence is also to be found in the correlation between the intelligence of children and that of their parents.

Pressey applied intelligence tests over 500 children and compared the result. He said that the scores of those children were much superior whose parents were engaged in professional or executive work than to those whose parents were manual labourers.

In India systematic study was carried to find the relation between the social status of fathers and the I.Qs. of children. The occupations of fathers were divided into six classes and a contingency table was prepared as follows :—

I.Qs.

Occupations of Fathers	Below 80	80–89	90–99	100-109	110–119	120 and above	Total	M I. Q.
1. Professors, Collectors, Deputy Collectors, Judges, Engineers and Asstt. Engineers.	1	3	6	6	9	17	42	114
2. Pleaders, Secondary School Teachers, Doctors, middle grade officers etc.	11	27	58	77	56	51	280	106
3. Primary School Teachels, Clerks etc.	32	37	53	62	43	31	258	100·5
4. Merchants, Landlords, Skilled Labourers, Typists.	27	40	50	42	23	21	203	[illegible]
5 . Petty traders, Farmers, Compounders, Priests etc.	31	33	31	23	13	10	141	93
6. Postmen, Policemen, Barbers, Mental Servants, Butchers etc.	40	13	13	4	5	1	76	81
Total	142	153	211	214	149	131	1000	

The coefficient of contingency was found to be 0·413 and the coefficient of correlation as 0·380 (by diagonal adding). This coefficient of correlation is fairly large. This shows that there is some unmistakable positive correlation, though not large, between father's occupational status and childrens intelligence. The difference between the mean I.Q. of the highest and lowest class is 33 points.

The study of twins made by Cyril Burt reveals that intelligence is more dependent upon heredity than on intelligence. The following table is indicative of the results of the study.

Correlation between tests of mental, scholastic and physical measurements.[3]

MEASUREMENT	A—Burt and Conway						B—Newman Freeman and Holzinger		
	Identical twins reared together	Identical twins reared apart	Non identical twins reared together	Siblings reared together	Siblings reared apart	Unrelated children reared together	Identical twins reared together	Identical twins reared apart	Non identical twins reared together
Mental (intelligence)									
Intelligence :									
Group test	·944	·771	·542	·515	·441	·281	·922	·727	·621
Individual test	·921	·843	·526	·491	·463	·252	·910	·670	·640
Final assessment	·952	·876	·551	·538	·517	·269	—	—	—
Scholastic									
General attainments	·898	·681	·831	·814	·526	·535	·955	·507	·823
Reading & Spelling	·944	·647	·915	·853	·490	·548	—	—	—
Arithmetic	·862	·723	·748	·769	·563	·476	—	—	—
Physical									
Height	·957	·951	·472	503	·536	—·069	·981	·969	·930
Weight	·932	·897	·586	·568	·427	·243	973	·886	·900
Head length	·963	·959	·495	·481	·536	·116	·910	·917	·691
Head breadth	·978	·962	·541	·507	·472	·082	·908	·880	·654
Eye colour	1·000	1·000	·516	·553	·504	·104	—	—	—

---

[3] The British Journal of Educational Psychology, November, 1955 Article—*Evidence for the Concept of Intelligence*, by Sir Cyril Burt.

Cyril Burt obtained these correlations from surveys in the London schools and supplemented them by further data collected by Miss Conway, who had been responsible for the final computations. The correlation obtained by Newman, Freeman and Holzinger have also been included here. "As regards intelligence the outstanding feature of the table is the high correlation between the assessments for identical twins even when they have been reared apart. It is almost as high as the correlations between two successive testings for the same individuals. Between non-identical twins the resemblances (at any rate with our own data) are not much closer than those between ordinary brothers and sisters. Nevertheless, environment is not entirely without effect, particularly when the assessments have been obtained by written tests applied to whole group. The effect is obvious when we compare the correlations for children reared together and children reared apart. And it might be thought that in the correlations obtained from unrelated children reared in the same homes we have a direct indication of its actual amount. In all probability, however, such correlations mainly reflect the method of placement : a dull or defective orphan would not be boarded out with a highly intellectual family. The figures for physical measurements, at least in our data, show very similar trends : with the American data the correlations are somewhat higher, but the disparity is seldom large.

The results obtained for the scholastic tests, both in the American enquiry and our own, present a striking contrast. In our own enquiry the correlations for siblings and non-identical twins reared together are actually higher than those for the identical twins who have been reared apart. And it may be instructive to note that the correlations which are most conspicuously increased by similarity of home environment are those for verbal or literary attainments ; those for arithmetical attainments are, if anything, increased more by similarity of constitution."

Figures like the foregoing provide ample evidence that the individual differences in general intelligence are in part at least inherited, and that they are affected by environmental differences much less than are school attainments.

There are, then, many indications that intelligence depends on heredity, and no evidence whatever that it depends solely or mainly on environments.

Intelligence and environments

There are at present two schools—the hereditarians and the environmentalists. If the I. Qs. of twins or siblings are close together the hereditarians say this is due to the same heredity, while the environmentalists aver that it is due to the fact that the children were brought up in the same environment. If the mean I. Q. of children in a lower stratum of society is less than the mean I. Q. of children in a higher stratum, the environmentalists say this is undoubtedly the result of different environment and the hereditarian says it may be due to different heredity since people of lower intelligence are driven by competition to the lower occupations.

The question of the constancy of I. Q. involves the important problem of the differential effect of various environmental conditions. Wellman examined children at succeeding ages and correlated the scores during the pre-school years with the scores obtained on the same individuals when they were in high school and colleges. Of 78 students, who were given college qualifying examinations 21 had attended a pre-school and 57 did not attend. It was found that the persons who had pre-school education made higher scores on qualifying and college examination. Thus pre-school attendance affected mental ability. A number of significant intellectual changes occured in many of the children ; one child changed from an I. Q. of 98 to 153 and another from 89 to 149. Wellman concluded that under ordinary circumstances change in I.Q. is very little but under various types of stimulation, however change may be tremendous.

Good-Enough has criticized this view. He says that Wellman made a serious error in her assumption regarding testing methods. The error involves the assumption that the kinds of abilities which are measured at the pre-school level are similar to those which are measured at high school and college level.

Anderson also criticized Wellman's conclusions. He recognized the effect of the environment both upon accelerating and decelerating the mental growth of an individual.

He pointed out that in many cases variations are found from one test period to another not because of intrinsic variations in intelligence but rather because of variability of achievement. He agreed with others that error in prediction are often due to common belief that the intelligence tests administered to a child of three are similar to the tests for a child of twelve. When a child is retested after a number of years, the disturbing factors may enter and chance failure and successes as well as varying opportunities for learning may seriously effect the resulting scores.

Most investigators have found a significant difference between the intelligence scores of children living in rural areas and those living in urban centres. It was formerly believed that this difference is due to selective migration from country to city. Pintner, for example, believed that the factor of selective migration to the cities of persons of superior intelligence accounts for this difference. Estabrook believed that the energetic individual does not remain in an area which does not afford him adequate opportunity for development. The implication is that the more intelligent individual is more energetic and thus is more likely to move to an environment allowing him a greater opportunity. The findings on carefully sampled populations contradicted the inferential evidence which has been employed to show that the more intelligent migrate to larger cities. Studies of Negroes have shown that those who live in cities attain higher intelligence ratings than those who have just migrated from rural areas. If migration is due to a selective factor, then the length of time which a person spends in a city should not be an important factor in making for difference in intelligence scores. Klineberg, in evaluating the problem of the relative importance of the racial heredity of Negroes and their environmental backgrounds, showed that the nature of the environment affected their intelligence scores.

Binet I. Q and Length of New-York Residence (Klingberg)

Group	Number of cases	Average I. Q.
Less than one year	42	81·4
One to two years	40	84·2
Two to three years	40	84·5
Three to four years	46	88·5
More than four years	47	87·4
New-York born	99	87·3

A sample of Negroes who lived in New York less than one year showed an average I. Q. of 81·4, and those who lived in New York for four years or more, an average I. Q. of 87·4.

The economic level and education of parents have been considered as important criteria employed in the determinations of the effectiveness of an environment in stimulating a child. One of the most important studies in this field was made by Free man and his associates. They studied that intellectual development of children placed in foster homes of varying socio-economic levels. They tested children before placement and rested them later in their foster homes. The children who were placed in better homes gained on the average of 5·3 I.Q. points where as those in poorer homes gained on the average of only ·1 point. They also found that when young children were placed in good foster homes they increased their I.Qs. much greater than other children. The correlation between the intelligence of separated siblings was lower than for the siblings reared together. This implies that the high degree of resemblance between siblings is to a large degree due to influence of the same environment.

The complex nature of the problem of environmental influence upon intelligence does not allow for specific conclusions. It is obvious, however, that we can no longer take for granted that heredity is the most important determining factor in the manifestation of intelligence. The evidence is clear that the nature of the environment in which a child lives influences his intellectual achievement.

S.C.E.R.T., West Bengal Library Calcutta 25/3, B. C. Road

CHAPTER 8

PSYCHOLOGY OF APTITUDE AND APTITUDE TESTS

What is an aptitude

WHAT PROGRESSION SHALL SUIT ME BEST ? WHAT TRADE should I undertake ? Should I go for a technical course ?

Insistent questions ! They are asked by young people with years of preparation ahead, and by mature workers grouping for secure foot-hold in an unstable society. They are asked not only of teachers and Heads in school or college, but of placement officers and employment officers, of organizers and social workers. These, as well as professional vocational councellors and consulting psychologists, are often called upon to confer with some one in doubt regarding his aptitudes and their relation to his future plans.

Now the question is as to what is an aptitude. When the psychologist speaks of an aptitude he has in mind the potentiality a person has to succeed in an occupation or school attainment. "Aptitude" is defined in Warren's Dictionary as "a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge, skill, or set of responses such as the ability to speak a language, to produce music, etc."

In referring to a person's aptitude for mathematics, or art, or carpentry, or law, we are looking to the future. His aptitude is, however, a present condition, a pattern of traits, deemed to be indicative of his potentialities.

In this definition nothing is said as to whether the 'condition or set of characteristic' is acquired or inborn. Too often it has been implied that the term 'aptitude' has reference to a person's native endowments only. Tests of aptitude, according to this mistaken view, should undertake to disclose the natural bent, the strength of the different dispositions, tendencies, and capacities inherent in the individual's original Constitution, without regard to the modifications in these capacities which have occurred in the course of experience. But this cannot be done. At least it is not possible with means at our disposal to-day. Even if it were, the resulting information would not be what is needed in individual counselling. We want the facts about a person's aptitudes as they are at present : characteristics now indicative of his future potentialities, whether he was born that way or acquired certain enduring dispositions in his earliest infancy, or matured under circumstances which have radically altered his original capacities is a question of great importance. An individual's potentialities at the period of his development are quite certainly the products of interaction between conditions both innate and environmental. His capacity for gaining manual skills, his intelligence, his emotional make-up, his moral character indeed all the aspects of his personality, are in varying degrees subject to limitations that have been imposed by opportunities for growth and experience, as well as by his original nature. No matter what his constitution may at first have been, it has unfolded, taken shape, been encouraged here and thwarted there, during the impact of favourable and unfavourable stimulation from the environments in which he has developed. And so, when appraising his aptitude, whether for leadership, for research, or for artistic design, we must take him as he is—not as he might have been.

Aptitude, moreover, connotes more than potential ability in performance, it implies fitness, suitability for the activities in question. How could one say that a young woman had an aptitude for nursing without first knowing not only her capacity to acquire the necessary proficiencies, but also the relative degree of satisfaction which the exercise of these proficiencies would yield ? One does not deem onself suited for a kind of work which would be strongly distasteful, nor even for an occupation so simple that it offers no challenge to one's interest. When appraising aptitudes, we are on the alert for symptoms of "ability to acquire" genuine absorption in the work, as well as a satisfactory level of competence. Indeed, a person who cannot develop a liking for an occupation along with proficiency in it, cannot properly be said to have an aptitude for it because he lacks the necessary drive.

Nature of aptitude

An individual's potentialities are not all equally strong. One can learn to do certain things more easily and better than other things, and can develop greater interest and satisfaction in some kinds of activity than in others. If you are a talented teacher possessing a professional efficiency it does not mean that you may necessarily be a good musician or an expert engineer. You cannot be equally superior in every trait.

Individuals differ one from another, in their potentialities. People do not all inherit the same endowments, nor develop equally. It was Sir Francis Galton who for the first time proved scientifically that there are individual differences. One individual may have the potentiality for becoming a good technician and other may prove to be a fool if, by chance, given this job.

The theory of aptitude implies that a person's potentialities are fairly stable. If he gives a clear evidence to-day of ability to learn to play the flute and of ability to learn accountancy, we should be amazed to find tomorrow that his aptitudes in these directions were reversed. Human nature is never as unstable as that.

Aptitude tests

There are various aptitude tests at present for a number of areas of performance, such as those involved in various occupations in the trades and industry, various broad areas of performance commonly dealt with in the school, and various narrow areas of performance largely unique to the school. The various types of aptitude tests largely possess in common the characteristic of testing the individual's potentialities in terms of specific abilities resulting from inheritance and general experience but of disregarding the abilities resulting from specific training or education. Thus aptitude tests parallel intelligence tests, although they are narrower in scope.

Achievement tests can also be used as aptitude tests. The difference between the achievement tests and aptitude tests lies, not in the tests themselves, but in the uses to which they are put. As a matter of fact many achievement tests are at the same time very good tests of aptitude. An achievement test in physics, a trade test in automobile valve-setting, and a proficiency test in the use of a slide rule are alike in measuring acquired abilities. They indicate how well certain things have been learned. But that is no bar to the use of person's scores in these very tests in estimating his aptitudes for learning something more. When a young man scores high in each of these three achievement tests—physics, valve-setting, and manipulation of the slide rule—one is justified in hazarding the opinion that he has some aptitude for engineering work.

Clerical Aptitude Tests

Tests to measure aptitudes for clerical occupations are of two sorts : general and specific. The general tests resemble the common verbal intelligence tests. Specific tests, on the other hand, measure particular abilities deemed symptomatic of capacity to learn to do one or more kinds of clerical work such as filing, stenography, proof-reading, or accounting. Some tests bring together in a single examination tasks to measure abilities both general and specific.

There are various clerical aptitude tests available these days. The Scott Company File Clerk's Test, a fifteen-minute examination which has been used for years as a general clerical test as well as for its original purpose, consists of six sub-tests : number-checking, name checking, alphabetical filing, numerical filing, topical filing, and phragraph comprehension. The clerical test of the British National Institute of Industrial Psychology, revised for use in England by Herbert Moore, furnishes another example of an examination designed to sample a person's abilities in each of several typical kinds of tasks. It is in seven parts, appropriately called Oral Instructions, Classification, Arithmetic, Copying

Checking, Filing, and Problems. The twenty problems in this last part are such as often appear in an intelligence examination.

Minnesota Vocational Test for clerical workers[1]

Among the well-known Clerical Aptitude tests is the Minnesota Vocational Test for clerical workers, adapted for either group or individual administration. This consists of number comparison and name-comparison tests. The abilities it measures are known to be indicative of certain of the aptitudes needed in a wide range of clerical occupations such as routine card sorting, straight typing, stenography, filing, cashiering, computing, book-keeping, and accountancy. The reliability of this test is satisfactory. It may be used at any educational level from secondary school to university. The complete test—known as the Long Form—can be given to an individual or to a group in about thirty-five or forty minutes. The first half—known as the Short Form—takes about twenty-five minutes to give. Each half is in two parts. The first part consists of 200 pairs of numbers ; and the second of 200 pairs of names. The instructions on the blank read "......If the two names or the two numbers of a pair are exactly the same, make a check mark (✓) on the line between them ; if they are different, make no mark on that line."

Samples of pairs of numbers done correctly are given below :

79542———79524

5794367 ✓ 5794367

Then are given samples of pairs of names done correctly :

John C. Linder———John C. Lender

Investors Syndicate ✓ Investors Syndicate

A brief practice exercise is provided, to make certain that the instructions are understood.

The prescribed time limits for the Short Form are :

(1)	Number checking	8 minutes.
(2)	Name checking	7 minutes.
	Total working time	15 minutes.

---

[1] The distributors of this test are—*Psychological Corporation*, Fifth Avenue, New York.

For the Long Form the time limits are :

(1)	Number checking	8 minutes.
(2)	Name checking	7 minutes.
(3)	Number checking	7 minutes.
(4)	Name checking	6 minutes.
	Total working time	28 minutes.

Scoring. Correct items are those in which 'sames' have been checked and the 'differents' left blank. The score for each test is the number of items correct minus the number wrong. Subtracting twice the number wrong from the number attempted gives the same results.

Various forms of *norms* for this test have been provided. For a full treatment of these norms and their significance, see Andrew and Paterson.[2]

There are many other clerical aptitude tests available these days. A detailed description of these tests is beyond the scope of this book.

During the recent years many attempts have been made to identify symptoms of aptitude for the various professions. Much work has been done towards discovering the aptitude for teaching, law, engineering, nursing and accountancy. Let us confine our discussion here to aptitude for teaching.

Aptitude or the Teaching Profession

Before we construct an aptitude test for the teaching profession it occurs to our mind as to what are the qualities or the traits of an ideal teacher. George Herbert Palmer,[3] in his volume 'The Teacher', has given a specific description of the ideal teacher. He puts his finger on four traits without which a teacher is handicapped. He places first a sympathetic imagination—a limberness of mind which enables one to think in terms of the other person's limitations, interests and needs, an aptitude for what he calls "altruistic vicariousness". Next is the necessity for a broad background of knowledge. The third requisite he describes as the power to kindle interest, to invigorate the productive activity of pupils as only a stimulating person can do. Finally, he stresses the capacity for disinterestedness—indifference to praise and recognition.

---

[2] Dorothy M. Andrew and Donald G. Paterson : *Measured Characteristics of Clerical Workers,* Minneapolis : University of Minnesota Press.

[3] George Herbert Palmer : *The Teacher,* Boston : Houghton Mifflin Company, 1908, pp. 3-30.

If a young person is drawn toward teaching, how is he to find out whether he possesses the golden qualifications listed by Professor Palmer, or the capacity to acquire them ?

Surely the necessary intellectual talents are the least difficult to appraise. Past performance in studies already pursued tells whether one has a bent for learning. The person who in his elementary, junior and secondary high school has not excelled the majority of his class-mates in ease and accuracy of a acquisition need not trouble to inquire further as to his aptitude for teaching. Anyone who is to become a teacher needs an intellect capable of grasping not only the subject-matter and its place in the curriculum, but also the aims and processes of education. But assuming that the candidate is bright, that he learns readily and assimilates thoroughly what he studies, the question still remains as to the likelyhood that he can also teach others. This aptitude sometimes reveals itself when class-mates come to him for help. If he enjoys explaining difficult points, if he succeeds in finding ways of classifying obscurities so that his fellows increasingly turn to him, that fact alone suggests that he probably has some aptitude for teaching.

The personal and social abilities—particularly sympathetic imagination and absorption in his pupils, without which a teacher fails to stimulate, counsel and instruct—are sometimes contrasted with the intellectual abilities which he exhibits in his own learning and productive thinking. Aptitude for teaching is set over against aptitude for scholarship, and it is asserted that the former peculiarly necessary in elementary and secondary schools, is in college and professional schools of only incidental importance in comporison with intellectual grasp of the subject to be taught.

It is indeed true that as the educational leader is ascended, intellectual power becomes increasingly essential.

As in other callings, certainly regarding one's interest in and aptitude for teaching is enhanced by even a modest opportunity to participate in it. Instructors and Councelors help those students who are seriously thinking of fitting themselves to teach, not only by telling them more about the things teachers do, but also by giving them a chance to share in some of them—helping a backward class-mate ; taking charge when the instructor is called from the room ; marking papers ; preparing a laboratory demonstration etc.

A person's aptitude may also be revealed in answering the following types of questions :

(1) Are you genuinely interested in young people or children of the age you would be teaching ?

(2) Are you able to direct and discipline others ?

(3) Are you enthusiastic about the work teachers do ?

(4) Can you adapt yourself easily to new situations ? etc.

The professional schools for training teachers have as yet found no sure means of ascertaining in advance whether a candidate has or can acquire the traits described in the foregoing paragraphs, although marked defects and shortcomings are usually recognizable during an interview with an applicant for admission. His intellectual grasp and mental alertness can be fairly well estimated by reference to his previous school record and his performance in tests of scholastic aptitude, vocabulary, English usage, mathematics, and other academic subjects, supplemented when necessary by such specially devised tests of intelligence as the one known as the Coxe-Orleans Prognosis Test of Teaching Ability.[4] This test was designed by W. W. Coxe and J. S. Orleans to measure ability to learn the subject-matter taught in normal schools. The five parts test the candidates general information, knowledge of teaching methods and practices as acquired from his observations as a student, ability to learn the type of material included in professional books used in teacher-training courses, ability to understand educational reading matter, and ability to study and work out educational problems. The total time for administering is approximately three hours. Detailed directions and scoring key are supplied by the publishers.

---

[4] Distributor : World Bock Co., Yonkers, New York.

In England the major problem for the educationists and psychologists is to discover whether a pupil in the primary school has an aptitude for academic, abstract and bookish education or for practical, constructive, creative, and technical education. There the children are given suitable secondary education according to their abilities and aptitudes. Now the question is how far the aptitude for grammar and technical school subject is predicted. There is much evidence on the prediction of grammar school success. Peel and Rutter and Emmett and Wilmut showed that performance in School Certificate subjects was highly predicted by predictive tests of intelligence, English, and Arithmetic, taken five years previously.

Evidence to technical school aptitudes is not so extensive, but correlations have been obtained which are comparable with those for grammar school subjects. Peel and Lambart, Watts and Slater all confirm that aptitude for technical drawing, wood work, and metal work can be precicted by tests. Work carried out at Birmingham University shows that aptitudes for engineering drawing, practical geometry and building subjects can be similarly predicted.

The aptitude test is only said to be predictive when the test scores correlate very significantly with the later measure of attainment or efficiency in the course of study or profession. This measure is called the 'Criterion' and the correlation between aptitude tests and the 'Criterion' gives the 'predictive power' of the test.

CHAPTER 9

ATTITUDE AND ATTITUDE TESTS

THE TERM 'ATTITUDE' HAS BEEN USED BY PSYCHOLOGISTS IN a great many senses, and there is no agreed definition. But it generally implies a personality disposition or drive which determines behaviour towards, or opinions and beliefs about, a certain type of person, object, situation, institution or concept. Often our attitudes are adopted ready-made, as it were, from our parents, teachers, or friends, though usually modified by our own experiences.

Development of attitudes in the individual

Environmental modification of the human organism begins with conception. Apart from embryonic development, the changes wrought by environmental contact are manifested from birth on during the early years or even months of life. Here the individual acquires his first attitudes and exhibits his first learned behaviour. In brief, he develops personality. Original drives are overlaid by social experiences organized into attitudinal systems.

The newly born child quickly acquires certain feeding habits and comes to regard favourably those who contribute to the gratification of his needs and unfavourably those who frustrate his activities. The baby's attitude toward his parents will carry over to others. As the child grows in his capacity to respond to those outside his family group, his social attitudes, *e. g.*, cooperativeness, selfishness, dominance, and the like, will become definite.

As the child acquires 'human nature' which will equip him for life in a society of human beings, the area of his relevant experience expands and the attitudinal patterns become incorporated in his personality. Attitudes will be modified, through learning, in accordance with his own goals and drives. This means that he will acquire attitudes like those of his parents, his friends, and the other primary groups of which he is a member. As the individual grows older, secondary groups with which he is associated will elicit certain attitudes. In general, the closer the relationships between individual and others, the greater will be the potency of such relationships in the formation of attitudes.

The growth and development of attitudes involves the integration of numerous specific responses of a similar type. Such attitudes determine the entire adjustment of the individual. If a child is frustrated by his class-room teacher and believes that the teacher is discriminating against him, he will tend to react negatively to the teacher's future criticism. Accordingly, the term 'attitude' is merely a convenient way of referring to the preparedness that exists within the organism for some future activity. It is to be emphasized that such preparedness is neither automatic nor routine, but possesses cognitive and conative aspects differentiating it from habits and reflexes as commonly regarded. Furthermore, attitudes toward inanimate forces, institutions, and values, while culturally determined, vary in intensity between individuals and within the same individual at different times.

The baby grows into childhood, youth, and adulthood with attitudes as an important aspect of his learned behaviour. Attitudes are evolved from association with his family group, with children in his play and school group, and in general through social-psychological interaction. As an adult, occupation and the responsibilities of citizenship and parenthood are important in attitude development. It is inconceivable that cultural stability, language, and dependability in human affairs could exist without attitudes. Yet the fact that attitudes do change account for social innovation, social conflicts, and social changes.

Opinions and attitudes

Opinions are verbal statements hypothesized to be symptomatic of attitude. When individuals express opinions, their statements can be considered to reflect their position on a scale which represents all degrees from a very favourable attitude to a very unfavourable attitude. A single opinion may provide very limited evidence concerning a person's attitude, or it may supply enough to locate his position on a scale with considerable precision. When the opinion expressed represents an extreme view point, it is more likely to place him correctly than when it represents a position nearer the middle of the scale.

Opinions are not the reliable index of attitudes. A man's pro-or anti-semitism, for example, is exceedingly complex, and could not be assessed satisfactorily merely by asking him if he liked jews. Actually the majority of attitude tests sample verbal expressions only, and the extent to which they penetrate to the 'private' level is doubtful.

The criterion for validating attitude may be corresponding behaviour. When one measures validity on a verbal level, the connection with actions must be established. Behaviour is not necessarily a true index of attitude. Hyman's study on behaviour of people replying to questions of a survey indicates people do not necessarily give a true picture of their behaviour as to redemption of war bonds.

Methods of Preparing Attitude Scales

There are two 'classical' scaling techniques which have been extensively used in opinions and attitude research since the early 1930's—the method of equal-appearing intervals associated with the name of "L.L. Thurstone and E. G. Chave,[1] and the method of summated ratings, due to Rensis Likert[2] and co-workers.

The Thurstone and Chave Method of Attitude-Scale Construction

For Constructing an Attitude Scale, statements representing a range of opinions concerning some matter may be collected and arranged in order according to the extent to which they represent a favourable or unfavourable attitude. A person could then read these statements and select the one most nearly representative of his own attitude, and the position of the statement thus selected would represent a crude measure of person's attitude. The instrument developed by this procedure would be unsatisfactory in many respects. There would be no assurance that the statements actually represented the full range of opinion, and neither would one be sure that the statements were properly distributed over the range of opinion covered. Adequate criteria for the selection of the statements included, in the scale are needed, and psychologists who have worked on this problem have suggested a number of different criteria that might be used. The earliest systematic attempt to establish and use criteria for this purpose was due to the efforts of Thurstone and Chave.

---

[1] Thurstone, L.L. & Chave, E.G. : *The Measurement of Attitude*, Illinois, Chicago, University of Chicago Press, 1929.

[2] Likert, Rensis : *A Technique for Measurement of Attitude*, Archives of Psychology.

Their method of constructing a scale of attitudes toward any particular object, institution, or idea involves the following steps :

(1) Several diverse groups are asked to write opinions about the particular object, institution, or idea. These statements should be brief and should be such that they can be endorsed or rejected in accordance with the attitude of the individual. Each statement must also supply evidence of a person's attitude. Statements should be neither ambiguous nor double barreled. Irrelevant statements should be discouraged by giving clear directions to the individuals expressing opinion. It should be noted that at this stage the written statements should cover the total range of opinion that it is desired to measure. Let us for example, write a few opinions about the 'church'.

1. The church has no value in promoting moral education.
2. People go to the church because they are forced to do so.
3. The church is an institution which helps in promoting moral education.
4. The church is a place where people learn to love God and the human beings.

5. The church must be discarded immediately.

6. The church has always cheated society.

These above statements are only examples and can be improved. Many statements of this form can be written. A selection of these statements ultimately forms the attitude scale.

(2) The next step is to place each statement on a scale of equivalent units, according to its favourableness or unfavourableness. It is based on the psychological method of equal-appearing intervals. Thurston usually employs several hundred judges, though reasonably accurate results can be obtained with as few as twenty-five. The statements are typed on separate small slips. Each subject is given a set of statements and asked to sort them into eleven piles, the left most pile, No. 1, containing items considered most favourable, and the right-most pile, No. 11, the most unfavourable. The other piles are intermediate and should be approximately equally spaced in degree of favourableness, in each judge's opinion. It should be observed that the assumption is made that the sorting process does not involve the attitude of the sorter. Sorters who have fundamentally different attitudes should still be able to agree on whether one statement expresses a more or less favourable attitude than another.

(3) Now the table is prepared for each statement. Let us suppose there are 100 judges to sort the statements. Let us further suppose that for statement No. 6 there were five judges who placed it in pile 5, eight who placed it in pile 8, thirteen who placed it in pile 9, and so on. These frequencies are then accumulated from left to right in the lowest row on the table as shown below. The number 13 is obtained by adding 5 and 8, and 26 is obtained by adding 5, 8, and 13, and so forth.

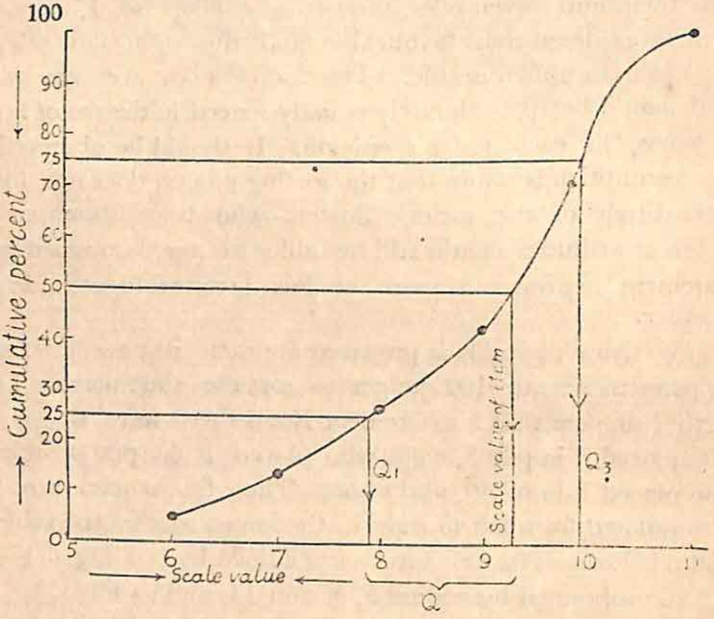
(*Data Concerning Statement No. 6*)

Scale value—

Q—

	Pile 1	Pile 2	Pile 3	Pile 4	Pile 5	Pile 6	Pile 7	Pile 8	Pile 9	Pile 10	Pile 11
Number of times item was placed in each pile						5	8	13	17	34	23
Cumulative frequency						5	13	26	43	77	100

A cumulative frequency graph is then plotted for each statement. The scale value for the item or the statement is the value which corresponds to a cumulative frequency of 50. In the case of the above statement, it would lie between 9 and 10 and is best determined graphically. This is done by plotting the data on a graph with the pile number on one axis and the cumulative frequency on the other, and the assumption is made that the pile number is the scale value for the mid-point of the pile. The graph shown below has been drawn from the data in the above table to show how this is done.



Graphical method of determining the scale value of the attitude item.

From the above graph it is clear that the scale value corresponding to a cumulative frequency of 50 is 9·2. This is called the scale-value of the item. The second thing we need to determine is the Q-value for each item. The Q-value is obtained from the graph by marking off the scale values which correspond to the 25 per cent (Q_1) and 75 per cent (Q_3) cumulative totals. In the above graph these values are 7·9 and 9·9 respectively. The difference between Q_3 and Q_1 represents Q-value. In the above graph Q-value is 2. Q-value is a measure of agreement between the judges. Q-value also indicates the extent to which the statement represents a clearly defined point of view. Statements with a high Q-values will be ambiguous or have other major defects. A high Q-value also indicates the disagreement among the judges. On a 9-point scale, Q-value should seldom exceed 2·0, and an average of less than 1·5 should be aimed at.[3]

A special problem arises in assigning scale values near the ends of the scale. If most of the judges assign a statement to pile 11, it may indicate that the statement might have been assigned by some subjects to a point beyond that pile if additional piles had been provided. In such cases the scale-value is estimated from the data in pile 10 and in piles with lower numbers. Consider the data shown below :

(Data Concerning Statement No. 4)

Scale value—

Q-value—

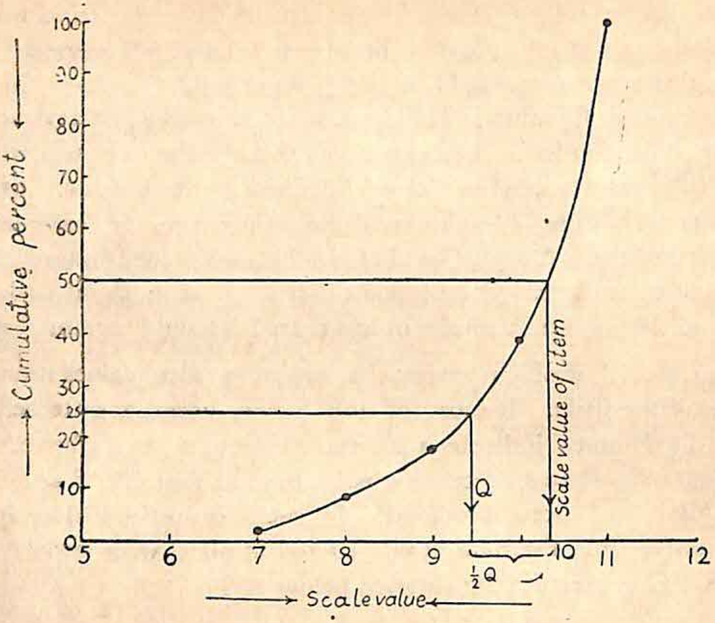
	Pile 1	Pile 2	Pile 3	Pile 4	Pile 5	Pile 6	Pile 7	Pile 8	Pile 9	Pile 10	Pile 11
Number of times item was placed in each pile	0	0	0	0	0	0	2	6	9	22	61
Cumulative frequency.							2	8	17	39	100

The data from piles 7, 8, 9, and 10 for this item would be plotted on a graph. The graph thus produced would be extraplotted to find the point corresponding to a cumulative frequency of 50 (50 per cent).

---

[3] Vernon, P.E. : *Personality Tests and Assessments* ; Methuen & Co. Ltd , London, 36 Essex Street, Strand, W. C. 2, 1953, p. 154.

This procedure is illustrated in the graph shown below :



Graphical method of determining scale-value of an attitude item

The value of Q in this case is found by taking the difference between the scale-values corresponding to cumulative frequencies of 25 and 50 and then doubling this value. In the above graph scale-value is 10·2 and Q-value is double of 1, *e.g.* 2.

The scale value and Q value of each item are calculated in the above way.

(4) The attitude scale is now built up by choosing a number of items (usually twenty or more) which have low Q-values, and whose scale-values cover the entire scale as evenly as possible. Thurstone and his students followed the practice of constructing two parallel forms of an attitude scale from the original large collection of opinion statements, each about twenty to twenty-two item long and each constructed as indicated above. The statements chosen may be presented on the final instrument in scrambled order as far as scale value is concerned. The attitude

scale is now ready for use. A subject whose attitude is to be measured is instructed to read each statement and check each one which he would endorse as expressing his own sentiment, opinion, or attitude. The subject's attitude score is then taken to be the mean or median of the scale values of the statements which he checked.

A typical Thurstone and Chave type of attitude scale is shown below. This scale was designed to measure a variable which might be named 'attitude toward teaching'.

(Scale of attitude toward teaching)[4]

Below are a number of statements about teaching as a profession. Consider each statement carefully and write 'yes' in the space before each statement with which you are in complete agreement and 'no' before all other statements.

Scale values

________—A good teacher should be an inspiration, a leader, and an adviser.

________—Teachers can teach parents to help their children.

________—The teaching profession does not need bright people entering it.

________—The social prestige of teachers should be the same as that of lawyers, doctors, and ministers.

________—There is opportunity to grow and learn in the teaching profession.

________—The routine nature of teaching makes it unattractive to imaginative people.

________—The main attraction of teaching is that it offers financial surity.

________—The financial rewards of teaching are a secondary matter in choosing it as a profession.

---

[4] Coombs, C. H. and Travers, R. M. W. : *Study of Attitudes Towards Teaching* ; Ann Arbor, Michigan University Press, 1947.

———————Teaching probably become monotonous after a few years.

———————Teaching is an occupation required relatively little skill.

———————The teacher plays a vital role in the development of good citizenship.

———————The work of teachers gives opportunity for expression— the work is creative, builds for the future.

———————A person whose main interest in teaching is that it provides a secure job should not be permitted to teach.

———————A teacher has to put up with many minor irritations in her work.

———————Students will inevitably have a high regard for good teachers.

———————Very able people would be wasting their time in teaching.

———————There are many personal rewards of the teaching profession—self-respect, personal worth professional competence, sense of creative power.

Score=Sum of weighted 'yes' answers divided by number of 'yes' answers.

Remmers[5] evolved a modification of the Thurstone technique of considerable practical value that makes it possible, with a single scale, to measure a very large number of attitudes. The discussion of this modified technique is beyond the scope of this book.

---

[5] Remmers, H. H.: *Generalized Attitude Scales—Studies in Social Psychological Measurements*, Purdue University, Studies in Higher Education: *Studies in Attitudes—A Contribution to Social Psychological Research Methods*, 1934, No. 26 : 7-17.

Likert's Method of Measuring the Attitudes.

In an attempt to save time and labour by doing away with the use of a judging group, Likert (1932), proposed that subjects be asked to express the 'extent' of their agreement with each of the items making up an attitude scale, instead of merely indicating agreement or disagreement in an all-or-none fashion. The terminology of the investigators who have utilized Likert's method varies to some extent, but in general it conforms to the following scheme : 'Strongly agree', 'agree', 'undecided', 'disagree', 'strongly disagree'. Arbitrary weights from 1 to 5 are assigned to these ratings in such a manner that either strong agreement with a statement that favours the institution in question or strong disagreement with one that is unfavourable to it is given a weighting of 5, while those at the opposite extreme are assigned a weighting of 1. Weightings from 2 to 4 are given to the intermediate judgments. The score of any individual is the sum of his ratings on all the items. Likert's method has appealed to many persons because of its apparent simplicity and because it enables the investigator to start directly with the group of subjects in whom he is interested without the necessity of preliminary standardization of his instrument beyond some form of item analysis used as a basis for the selection of the statements to be included in the scale.

Likert uses other types of items also. This type is illustrated by the following attitude-test item.

How far in the educational system (aside from trade education) should the most intelligent Negroes be allowed to go ?

(1) Grade School.
(2) Junior High School.
(3) High School.
(4) College.
(5) Graduate and Professional School.

Each question thus constitutes a short attitude scale in itself. This is an advantage, for it means that each item yields more information than do items to which there is only a Yes-No response.

Advocates of the Likert technique have commonly included a great number of items in their scale than have those utilizing the method of equal-appearing intervals. This means that more time will be needed for scoring the papers, a fact that to some extent it does away with the avantages of eliminating the judgment group. Likert's method assigns equal weight to all the items although the chances that all indicate equally favourable or unfavourable attitudes are slight.

Since the teacher will probably not be sufficiently familiar with the statistical techniques needed for developing a Likert-type scale, it will not be further elaborated upon here.

CHAPTER . 10

PERSONALITY AND ASSESSMENT OF PERSONALITY

Definitions of Personality

LIKE THE COMMON SENSE DEFINITIONS OF INTELLIGENCE, THE common-sense definitions of personality have had rather heavy going. So many different definitions have been offered that the whole enterprise has almost bogged down by its own weight. Allport for instance, lists some fifty definitions of personality that have been used at one time or another.[1] Many of these definitions, of course, are concerned with the legal or the theological aspects of personality and not with the psychological aspect. Those definitions of personality which are psychological in nature tend to stress one or more of three different features namely : (a) the over-all or inclusive nature of personality, (b) the social-stimulus aspect of personality, (c) the role of personality in making us unique or individual persons. At one place Allport defines personality as the dynamic organisation within the individual of those psychological systems that determine his unique adjustment to his environment.

Burt defines personality as 'a unified pattern of bodily and mental reactions exhibited by the self in response to a social situation'.

Munn describes personality as the most characteristic integration of an individual's structures, modes of behaviour, interests, attitudes, capacities, abilities, and aptitudes.

---

[1] Allport, G. W., '*Personality, a Psychological Interpretation*' ; Holt, 1937.

Many of the definitions of personality stress the feature of *uniqueness* or *distinctiveness*. In these definitions, personality is that quality or combination of qualities which sets us off from other people.

Operational or working definitions

The common sense definition of personality represent what we would like to know about this subject, or what we set out to know. These definitions probably describe the general concept we have in our minds when we begin to read a book or a chapter which professes to tell us about personality. It is most probable, too, that these were the general ideas in the minds of the early investigators who first set out to study the facts of personality. They undoubtedly hoped to get some information about the integrated, unique person in his relation to other people.

It is helpful to keep the hopes and intentions of the investigator in the back of our minds. It would be a great mistake, however, to assume that the investigator has always observed exactly what he set out to observe. With the very best of intentions, he may design tests which he hopes will reveal aspects of personality as it is commonly understood, only to find out later on that his tests are really picking up something very different.

If we can bring ourselves to accept this hard and unpleasant fact, we will realize that here, just as in the case of intelligence, the phenomenon studied is really, 'defined' by the nature of the tests or by the techniques of observation. Our knowledge is limited to the things that will come through these measuring instruments. If our tests can reveal the 'integration of traits', which make a person unique, then that is what we are actually studying.

If the tests or observational techniques, on the contrary, reveal only the ability to distinguish pretty objectives from those which are reprehensible, then that is the trait we are studying. Adopting this hard-boiled operational approach, we should be prepared to find personality as investigated to be somewhat different from personality—as envisaged-on-the-mountain top. To a great extent, the investigations do not stress integration or uniqueness. They stress, on the contrary, a large number of separate traits. For the most part, the investigations have used measures of such narrow traits as honesty, sociability, introversion, neuroticism, and the like. And it is this aspect of personality that we shall learn about from the investigations. The personality which we will study will very frequently turn out to be a cluster of traits in which people vary.

Personality Traits

In this chapter we are mainly concerned with attempts to measure or judge personality traits, and therefore we must be clear about this term. Personality traits can be regarded as permanent clusters or constellations of more specific situations. Thus we may talk of the traits of perseverance. If it is a permanent feature of a person's make-up, perseverance will be shown in many little specific situations—home-work, gardening, hobbies, dealings with other people. It is difficult to measure the personality traits. Firstly, they are mostly very vague and ambiguous in meaning, and different people often include different modes of behaviour within anyone trait. For example behaviour which one person interprets as aggressive might be called adventurous by another. Secondly, personality traits involves subjective interpretation. They are partly dependent upon the observer. His own personality and his view point both influence what he notices in other people's behaviour, and his interpretation of the traits are responsible for such behaviour. Thirdly, any specific behaviour depends on such a multiplicity of factors in the personality structure and the environment, that no one reacts in accordance with a trait all the time.

Does this mean that we should not try to assess the personality traits and if we do so our judgments are faulty. No we find much truth in rating the persons for various personal, emotional, and temperamental qualities, e.g., for narrow interests, willingness to cooperate, curiousity and joy etc. These ratings are correlated and groupings or clusters of allied traits are discovered by factor analysis.

Types of Personality

There is a difference between the traits and types of personality. Types are more general than traits and can be regarded as constellations of traits. Very often types are dichotomous, *e.g.* the introvert, *viz.*, extravert types. The extravert-type may include the traits of cheerfulness, shallowness, self assurance etc. while the introvert includes lack of self-confidence in the society perseverance, accuracy etc. Individuals vary in personality types. Carl Jung has divided human beings into two primary types, which are known as the 'extraverts' and the 'introverts'. According to Jung the extravert personality is interested in the outside world, in reality for its own sake. He is usually cheerful, self assured and talkative. He may be aggressive and demand approval, when in doubt or difficulty he turns to society for assurance and help. He has a low level of aspiration and shows no tendency to underestimate it. The introvert is turned into himself and views the world subjectively. He may lack assurance in company. He waits for approval and when in trouble rarely takes his worries to his fellows. He has high level of aspiration and severely underestimates his past performance. An example will distinguish between these two types. Let us suppose for instance, that your partner has absconded with your money. This produces severe tension or frustration. How can you release those tensions. If you are an extreme extravert, you will get most satisfaction from vigorous outward action. You may try to apprehend the thief. In some cases you may resort to fighting. After doing so you may feel some release. If you are an extreme introvert, on the other hand, you may get your greatest relief from rearranging your ideas to make such conduct understandable. After you have thus reorganised your philosophy of life, you may feel better about the problem. People do not thus fall neatly into two exclusive compartments, and Jung defined a third type of individuals, the 'ambiverts', who possess some of the qualities of both extraverts and introverts.

The opposition of two easily recognised forms of insanity, cyclic or manic-depressive and schizophrenic or dissociative insanity have led to one commonly used opposition, cyclothyme vs. schizothyme.

The Assessment of Personality

There are four main methods of assessing the personality :

(1) The interview.

(2) Ratings and judgment of personality.

(3) Personality questionnaires.

(4) Projective techniques.

(1) THE INTERVIEW

This is one of the oldest and most direct methods of assessing personality. For obtaining a rough idea of the personality of the interviewer one single formal session may be conducted. A better procedure is to conduct a series of interviews, interspread with observations of the interviewer in different situations. Interview seems to be a satisfactory method because it makes use of observations of external appearance, of gestures, voice and other modes of expression, and of behaviour under the stress of interview situation or in response to difficult questions. Interview is universally accepted both by the interviewers and the interviewees. It is quicker and economical than any method involving tests. It is flexible to new purposes to special circumstances.

Interview is not free from limitations. It provides an unrepresentative and limited sample of the interviewee's behaviour. He is keyed up to make a good impression, and subjected to a situation very different from that commonly occuring in his later school career or job.

Interview can be used with advantage for supplementing other methods of personality assessment.

(a) **Reliability of the interview**. For determining the reliability of the interview test-retest method can be used. But this method does not give a satisfactory reliability coefficient because the subjects' responses may differ when they meet a situation a second time and moreover personality qualities are admitted to be somewhat unstable. A satisfactory method of determining the reliability of interview is to see the agreement between two or more observers in recording specified types of behaviour. By this method the reliability co-efficient can be expected to approximate 1·0. In practice we generally get reliability co-efficient around ·6.

(**b**) **Validity of tha interview.** Vernon and Parry quote studies where particular psychologists or personnel selection officers (PSOs) were notably successful in picking recruits for service employments or for officer training. There are very few evidences regarding the high validity of the interview, [2]McClelland's study shows that teachers' judgments of industriousness or other personality qualities are not valid.

In conclusion it may be said that interview requires an expert training. An untrained interviewer can hardly assess the personality correctly. Interview cannot be discarded inspite of its certain limitations. It is desirable to develop more objective methods. Interview, if carried out in an objective way is very practicable and can be used for borderline cases in selection procedures.

(2) RATINGS AND JUDGMENTS OF PERSONALITY

Every day we rate other persons and are ourselves rated by them. As soon as we meet a person we jump to conclusion about him. We try to interpret his expressive movements, his feature, his actions etc. and thus arrive at a kind of picture or 'schema' of his personality as a whole. The main rating techniques are as below :

(a) Ranking and paired comparisons

(b) Numerical ratings

(c) Man-to-man scale

---

[2] McClelland, W. : *Selection for Secondary Education* ; Londc1 University of London Press. 1942.

(d) Verbal and graphic scale
(e) Analytic scales
(f) Voting and Guess-who Technique
(g) Ratings within persons.

(a) Ranking and paired comparisons. We try to rank the persons according to merit for a personality trait. But this method is incapable when the number of cases exceeds twenty.

In the paired comparison method the rater is given every possible pair of names and asked to say which of the two is higher. The results can be expressed finally as a ranking score.

(b) Numerical Ratings. Each person is given marks for traits. These marks can be graded in five or seven point scale e.g., 5, 4, 3, 2, 1 or 2, 1, 0, —1, —2 or turned into letters, A to E, or verbal labels : Excellent, very good, good, fair, week, and so on.

(c) Man-to-Man Scale. The man-to-man scale has proved useful both in rating a group of people on several traits at once. It was developed in the first world war in rating men for potential officer material. To use this scale, all the raters must know a number of people who can be lined up in an order, such as the following :

Individual	*Agreed Judgment of the Rater*
Ram	—Best man to go through this school
Mohan	—Superior
Suresh	—Average, good man
Kishore	—Barely passable
Vijai	—Not acceptable

once the raters have agreed on such a scale, they may rate new men in terms of these known men.

(d) Verbal and Graphic Scale. Instead of letters or numbers we can use terms like : Excellent, very good, good, fair, weak etc., for rating the personality traits. In the case of graphic scale each term is defined concretely so that the rater no longer has to think quantitively or bother about standards. A graphic form in which descriptive phrases are located along a straight line has been illustrated below :

Attends to Important Details

1	2	3	4	5
Undependable	Most of the time	About Average	Seldom neglects	Never neglects

The rater merely puts a tick or cross at whatever point on the line that he thinks appropriate.

(e) **Analytic Scales.** This method is breaking down a trait into a small number of 3-point sub-scales which are separately rated and the scores combined.

(f) **Voting and Guess-who Tachnique.** If children of 9 years age or so be asked to vote for some other children to whom they think highest or lowest in various respects, they can do so sufficiently accurately.

In Guess-who Technique series of short character sketches, e.g., very lazy, lazy, an average, active and very active, are drawn up. The class is asked to guess whom these represented. The following is an example :

These three people are always happy and enjoy themselves. It is impossible to annoy them. They never change. Guess who they are.

(1)——(2)——(3)——

(g) **Ratings within persons.** This refers to the rating or ranking of a number of traits according to their prominence within an individual, as contrasted with the ordinary procedures of rating a number of individuals on a trait. In some way it is easier to judge whether a person is most outstanding for, say, sociability, instability, etc. than to arrange numerous individuals according to degrees of sociability.

Limitations of Rating Scales

(1) It is very difficult to rate persons accurately. When we say that he seldom neglects work it does not convey any objective idea. Different raters may have different opinions on this.

(2) The rater is influenced by the 'halo' effect. A rater, that is to say, who thinks well of a student in one may tend to rate him highly in most other traits as well. If Kishore is your favourite student you will naturally be inclined to rate him favourably in sociability and industriousness etc.

(3) The bias of the rater is more apparent sometimes than others. In rating any trait there is a certain reflection of the rater's own taste.

Devices for Improving Ratings

(1) It must be noted that relative ratings are always superior to absolute ratings.

(2) Graphic scales are better than numerical or letter scales.

(3) Analytic scales should be used whenever possible.

(4) Each trait must be defined concretely before rating any person for that trait. So the trait of like ability may be defined as follows. 'A likeable child is one you would take on holiday with you'. A 'likeable child is one you would choose to represent the school at a scouts jamboree.'

(5) When several traits are to be rated, rate all individuals on one trait at a time, not each individual on all traits.

(6) When an individual is to be rated for numerous traits then it is a good policy to arrange the items on the form so that the relatively desirable and undesirable extremes alternate in random fashion. Otherwise the rater is liable to go down the page checking the desirable (or undesirable) answers or grades throughout.

(8) Independent judgments of three or four raters should be taken into consideration for cancelling out individual prejudices.

(9) Raters must be given a training in the use of rating scales.

(10) Raters should have had plenty of opportunity to observe kind of behaviour they are rating.

(3) PERSONALITY QUESTIONNAIRES

Most of the data on personal adjustment have been gathered through some form of personal data sheet, personality schedule, or personal inventory. Typically these questionnaire contain a long list of symptoms or complaints often reported by people who have been classified as neurotic or extravert or introvert and so on. The person taking the test then merely indicates whether or not he has ever experienced this symptom.

Most of the hundred or more tests that have been published are modifications or extensions of three prototypes : 'Woodworth's personal data sheet', 'Freyd-Heidbreders Introversion Extraversion Test' and 'Allports' A—S (Ascendence Submission) Reaction Study'.

Personal Data Sheet

The 116 items in [3]Woodworth's 'personal data sheet' were originally derived from medical psychologist's descriptions as symptoms of neurotic patients. The following are some representative examples :

(1) Do you usually feel well and strong ?

(2) Do you ever walk in your sleep ?

(3) Do you usually feel tired upon awakening in the morning ?

(4) Have any of your family had a drug habit ?

(5) etc.

Each question is followed by 'yes', 'no', and one of this is to be checked. The person taking the test then merely indicates whether or not he has ever experienced any of these symptoms. A person who checked a large number of the complaints typical of neurotics would obtain a low score on personal adjustment.

---

[3] Woodworth, R. S. : *Personal Data Sheet* ; Chicago, Stoeltin, 1920.

Lairds' personal inventory B-2 [4]contains similar items but with multiple-choice responses, Thurstone's personality schedule test[5] contains 223 items collected from Woodworth, Laird and other sources. Percentile norms are available for college students.

Freyd-Heidbreder's Introversion and Extraversion Test[6]

Freyd collected 54 items descriptive of the introvert type from Jung's writings, of which the following are example :

(1) Is self-conscious

(2) Is slow in movement

(3) Keeps in the background on social occasions.

(4) Prefers to read a thing than experience it.

Heidbreder turned these into a self-rating test, where the tester checks each item +, ? or −

Allport's A—S Reaction Study[7]

Here the items, were made up to represent concrete, manifestations of dominatingness (ascendence or submissiveness, and were standardized by comparing the answers of students who had been rated by associations as highly ascendant or submissive. The following is an example : The numbers show the weighted scores for ascendence.

You arrive late in the church, and all but the front seats are full. Will you go down to the front of the church ?

---

[4] Laird, D A. : *Detecting Abnormal Behaviour* ; J. Abn. Soc. Psychol., 1925, pp. 128–141.

[5] Thurstone, L. L., and Thurston, T. G. : *A Neurotic Inventory* , J. Soc. Psychol., 1930, 13–32.

[6] Freyd, M. : *Introverts and Extraverts* ; Psychol. Rev., 1924, 31, pp. 74-87. Heidbreder, E. *Measuring Introversion and Extraversion* ; J. Abn. Soc. Psychol., 1926, 21, pp. 120-134.

[7] Allport, G.W. : *A Test for Ascendence – Submission* ; J. Abn. Soc. Psychol. 1928, pp. 23, 118-136. Allport, F. H. and Allport, G. W. ; *A—S Reaction Study* ; Boston, Houghton Mifflin, 1928.

Always..............3

Sometimes...........0

Never.............—3

An alternative form is available for women, and an adaptation for children has been prepared.

Tests of other traits

There are many tests of other traits but Hathaway and McKinley's Minnesota Multiphasic Personality Inventory (MMPI)[8] and Bernreuter's Personality Inventory[9] are better known.

Minnesota Multiphasic Personality Inventory

This test is widely used in mental hospitals of England and America. There are 550 statements. They are generally presented individually, on separate cards and sorted by the patients into 'True', 'False' and 'cannot say' boxes. It takes from 30 minutes to several hours to complete. On the basis of the responses of 500 normal adults (16 to 55 years) and 800 miscellaneous patients, a series of empirical scoring keys has been developed, so that a profile is obtained showing relative scores on : Depression, Hysteria, Masculinity, Feminity etc.

The Bernreuter Personality Inventory. This inventory measures four traits—Neurotic tendency, Introversion—Extraversion, Dominance—submissiveness, and Self-sufficiency. Each of its 125 items is scored for each trait. For example the responses to : Do you day-dream frequently are scored :

	Neurotic	Introversion	Dominance	Self-sufficiency
Yes	+5	+3	−1	+1
No	−4	−4	−1	−1
Doubtful	−2	0	+2	+2

---

[8] Hathaway, S.R., and Mckinley, J.C. : *A Multiphasic Personality Schedule (Minnesota)*, J. Psychol., 1940, 10, pp. 249-254.

[9] Bernreuter, R.G. : *The Theory and Construction of the Personality Inventory* ; J. Soc. Psychol. 1933, 4, pp. 387-405.

Questionnaires for children

It is difficult to rely upon the answers of children (under 14 years) to personal questions. Maller constructed a questionnaire for children which is known as Maller's Character Sketches. It contains 200 short descriptions ; the testee has to say whether or not he feels or acts like the person described. For example :

—This person finds it difficult to forget unpleasant memories and cannot help thinking about them.

Each item is repeated in reverse form, elsewhere in the test, as a check on consistency, e.g.,

—This person does not find it difficult to forget unpleasant memories and may not think about them.

Limitations of Questionnaire

1. The person taking the test may 'see through' the questions and check the answer which he thinks will give him a 'good' score rather than the answer which he regards as true. When the testee is aware of the traits for which the questionnaire is meant he tries to deceive others or he may even deceive himself. Although the name of the trait or traits at which a test is aimed is usually withheld, and some non-committal title like inventory or A-6 study is given, it is obvious that testees will make their own guesses as to the object of the test, and will answer each question not so much at its face value as in accordance with their own interpretation of the object, and with how much they are willing to reveal.

2. Sometimes the testee answers the items of the questionnaire in a particular way not because of his weakness in some personal adjustment but because of some physical disability. So if a testee replies to Allport's A—S Reaction study item, regarding the church, by saying 'Always' it does not necessarily mean that he is ascendant. It is possible that he goes down to the front seat not because he is ascendant but because he is hard of hearing. In such cases questionnaires give a false index of the trait.

3. The influence of temporary mood-optimism, worry, etc. might also be thought to affect test responses. It is sometimes difficult to devise a test situation which is a natural stimulus for temperamental responses.

4. The empirical validity of the questionnaires has also been the subject of much discussion.

For determining the empirical validity of the questionnaires the psychologist correlate the scores with other evidences of the traits. The study of Ellis shows that the validity coefficient of the questionnaire is ·4. This validity coefficient is too low for making predictions about the individual.

It may be concluded that, despite their extreme weaknesses and dangers, questionnaires should not be entirely condemned. Well constructed one's given under suitable motivating conditions, can be of value both for experimental research, and in clinical or other applied psychological work.

4. PROJECTIVE TECHNIQUES

Some persons say that projective techniques have been originated from the Freudian theory of 'projection'. The theory is when our sense of self-respect is assailed by some unacceptable drive, we may readily project that drive on to someone else. We may accuse that other person of the sin which tempts us. Projection in its many forms guards our sense of self-respect in two ways. It may actually transfer the guilt to someone else. We may be completely convinced that all the hostility resides in our colleague. And thus we are completely relieved from the feeling of guilt. Projection is seen when people blame the tools or circumstances for their failure. In this form it is very common in children. 'It was the pen that did it.' 'The table is too high'.

In the second case projection may fail to achieve a complete transfer of guilt. We may be vaguely aware of our hostility, but because of our belief in the hostility of the other one, we may feel better about our own half conscious aggression. It may also make us feel better about our half conscious guilt.

Few of the numerous tests which are commonly grouped under this heading directly involve Freudian projection ; rather they provide a vehicle through which the subject expresses his personality structure.

Frank[10] has pointed out that they originate from psychoanalytic techniques of dream analysis and free association on the one hand, and from Gestalt Psychology, with its emphasis on the whole as more than the sum of its parts, on the other. The dream is 'the royal road to the unconscious'; and 'the creative productions of writers and artists' have often been claimed by psychoanalysts to reveal underlying personality trends and conflicts. At the same time, no single trait, mechanism, or complex manifests itself in isolation from the rest of the personality structure.

There are many projective techniques that are used these days but let us discuss only those which are best established.

(1) Free Word Association

This technique was developed by Galton in 1879. These days it is not much used. In this technique usually a list of 50 to 100 stimulus words is read out by the tester; to each word the subject responds with the first word that comes to mind. Many associations are superficial verbal habits—opposites, rhymes, etc., e.g., black-white, father-mother, etc. These hardly help us to discover any trait. But a few stimuli may touch on emotional complexes and lead to personal responses. The psychologists are able to interpret the responses of the subject to discover his personality trait. The tester classifies the responses of the subjects but Wells and Murphy show that there are considerable discrepancies when the same responses are classified by different testers, and that there seems to be no correlation between the types to which a subject is prone and his personality traits or his neurotic or psychotic syndrome[11].

Pressey Cross-out Test[12]

This test was constructed by S.L. Pressey in 1919 which does not measure any particular trait but reveals the stimuli that evoke emotional responses. It is in two forms : 'Form A' for adults and 'Form B' for children. Form A contains four sub-tests each with 25 rows of five words. In the first sub-test the subject crosses out all words that are unpleasant to him and encircles the most unpleasant in each row. In the second, each set of five is preceded by a word in capitals : the subject crosses out the words associated with this stimulus and encircles the most closely associated. In the third, 'things regarded as wrong are crossed out and encircled and in the fourth, things about which the subject has worried or felt nervous. Form B contains three similar sub-tests, calling for reactions of wrong, worry and like or interest, respectively. Pressey regarded the total words crossed out as a measure of 'affectivity' or 'richness in emotional associations'.

---

[10] Frank, L. K.: *Projective Methods for the Study of Personality*; J. Psychol., 1938, 8, pp. 389–413.

[11] Wells, F. L.: *Association Type and Personality*'; Psychology 1919, 26, pp. 371–376. Murphy, G.: *Types of Word Association in Dementia Praecox, Manic-Depressives, and Normal Persons*; Amer. G. Psychiat, 1923, 2, pp. 539–571.

[12] Pressey, S. L.: *A Group Scale for Investigating the Emotions*; J. Abn. Soc. Psychol., 1921, 16, pp. 55–64; Test published by Stoelting, Chicago 1919.

Incomplete Sentences Tests

One of the limitations of free word association is that the single-word responses provide little material for analysis. A. F. Payne constructed the 'Sentence Completions' test in 1928 which meets the above limitation. In this test some short incomplete phrases are presented in the printed form, such as :

My brother...............

Home is...................etc.

The subject is asked to complete each of the phrases by writing a few words that occur to him first. This test reveals much of the emotionally toned material. This test can be used as a group test.

TAT

Among the most prominent projection techniques are the Thematic Apperception Test (T.A.T.) and the Rorschach Ink Blot Technique.

Thematic Apperception Test

This test was constructed by Morgan and Murray[13]. In this method, the subject is shown a picture, for instance, of an older man taking hold of the arm of a startled young woman. The subject is told that this is a test of fertility of the imagination. He is asked to give a story into which the picture would fit. The resulting story is supposed to give some insight into the hopes and fears and needs of the subject. A special group of such designs have been provided for adolescents (17), and still other forms have been devised for young children. One of these, 'the childrens' Apperception Test,' makes considerable use of animals on the assumption that children might more readily project their needs and fears into animals. So far the evidence suggests that the use of animals adds very little and that elementary school children give no different response to these than to the standard tests.

---

[13] Morgan, C. D., and Murray, H. A. : *A Method for Investigating Fantasies : The Thematic Apperception Test* ; Arch. Neurol and Psychiat. 1935, 34, pp. 289-306 ; Test Published by Harvard University Press, 1943.

Morgan and Murray collected a series of 30 photographs (10 for women, 10 for men and 10 for both). These photographs are used as described above. It does not take more than an hour to show the photographs to the subject and get his responses. Even if the subject is not able to cover all the photographs time limit does not exceed an hour. A second interview may be held, or it may be used for discussing the previous stories and asking the subject himself to help in interpretation.

Generally a set of 10 photographs is used. But each of them should portray an individual of the same sex and about the same age as, or younger than, the subject, so that he can readily identify himself and project his own needs and sentiments, frustrations and conflicts, etc. This test can also be used as a group test in which case the photographs are projected by slides for a few minutes each and each individual is asked to make up a story describing the situation.

An interesting continental version of T.A.T. is Vann Lenneps, '4—Picture test'. [14]Jackson's set of 6 pictures, 'A test of Family

---

[14] Jackson, L. : *Emotional Attitude Towards the Family of Normal, Neurotic, and Delinqueat Children* ; Brit. J. Psychol. 1950, 41, 35-51, 173-185 ; Test published by Methuen, London, 1952.

Attitudes', appears suitable for use with children in England. In India it can be used after adaptation.

The interpretations of T.A.T. differs with the theoritical background of the psychologists. According to Murray the needs and Presses (external forces that the individual regards as beneficial or harmful) of the individual play the most prominent part in the make up of the story. Symonds claims that the stories mainly represent pressed drives which may be very reverse of the overt personality traits. Burt and sen in their unpublished memorandum advocate a still more straight forward approach, owing little or nothing to abnormal psychology. They assess the strength of a number of qualities such as level of Organization, Observation of Details, Verbal Richness, Imagination, Extraversion-Introversion of the main themes, Maturity vs. Childishness, Integration vs. Neuroticism.

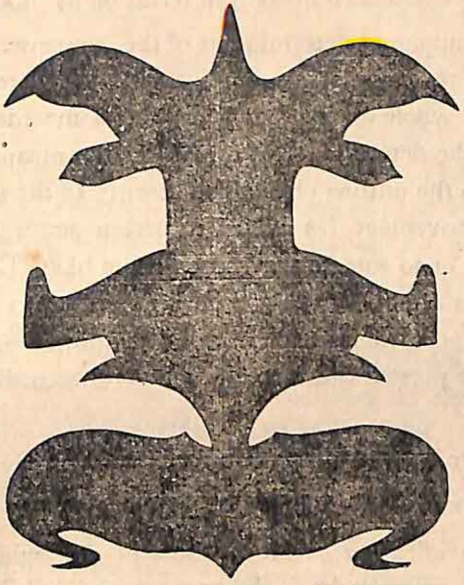
T.A.T. if used along with other tests like Rorschach Technique etc., proves to be useful for mental hospitals and clinics, etc. There is evidence to show that interpretation of the same material by different testers are reasonably consistent, and have achieved a high degree of success in matching interpretations with independent case histories.

The Rorschach Ink-blot Test[15]

The best developed projective technique which makes use of a highly unstructured situation is the Rorschach Inkblot Test, which is of much greater interests to the clinician than to the educator. The Rorschach technique was originally developed by a Swiss Psychiatrist by whose name the test is known. It consists of a sequence of ten cards, each one of which presents an ink-blot which may be in black and white or have colours in addition. The blots themselves were originally selected by Hermann Rorschach himself from thousands which he tried out The significance of the series of blots lies not only in the fact that each blot has been selected on the basis of sound reasons, but also because the blots have been selected to form a definite sequence. They should not be used in just any sequence, since it is important, for example, that card 2 follow card 1 and not the reverse.

---

[15] Rorschach, H. : *Psychodiagnostics* ; New York, Grune and Stratton, 1942, p. 226.



A sample card of Rorschach Test.

One of the Rorschach's major discoveries was that features other than content are important in interpreting what a person sees in a blot. Psychologists prior to Rorschach had made extensive analysis of the responses of the individuals to ink-blots, but the emphasis was on 'what' they saw. Rorschach realized that the most significant aspects of a response might be discovered by a study of factors other than content.

Administration of Rorschach Test is simple. The blots (some black-grey-white, some coloured) are presented in turn, and the subject is asked what he sees in each, what it makes him think of. He is encouraged to produce as many associations as possible, and to turn the cards round if he wishes. After completing the series, the responses are discussed in order to elucidate what determined them, and to which part of the blots they refer.

Rorschach developed a system of classifying the responses according to what he considered were the most important features. The basic essential of Rorschach system is to give each response a two-fold classification in terms of its 'location' and in terms of the apparent determinants of the response. By location is meant whether a response names some object represented by the blot, as a whole or whether it refers to some major or minor detail. By the determinant of a response is meant whether the response is to the outline of the blot (form), to the colour, to the shading, to movement (as when a person states that he sees a running dog) or to some other aspect of the blot. Certain categories of content are also given special emphasis in the interpretation of the report, such as the meaning of human or animal objects. At the present time there are several recording systems in use.

The above discussion can be summarized briefly as below.

Each response is scored under three headings :

1. *Mode of appreciation*. Whether the whole blot (W response), an ordinary detail (D) or unusual detail (Dd)', or based on a white space in the blot (DS), etc.

2. *The determinant*. The form or shape (F+or−according to aptness), colour (C), shading or chiaroscuro (K), or a response implying movement (M). The relative proportions of colour and movement responses yield the Erlebnistypus or experience balance—extratensive (excess of colour), introversive (excess movement), many of both types (detailed ambiequal), few or more (Constricted).

3. *Content*. original (o or —), or common response (P) ; human (H), animal (A), anatomical, object, geographical nature, architecture, plant, etc.

For a normal production of about thirty responses testing usually takes less than half an hour but there are enormous variations ; some patients give less than ten responses, some several hundreds. Scoring and interpretation may take several hours.

Unfortunately, Rorschach did not live long enough to see his work acclaimed by the world as a major development in psychological assessment, but he must be credited with having evolved every major idea on which present Rorschach technique is based. The subsequent work of Beck, Hert, Klopfer, and others must be considered as that of refining the techniques developed by Rorschach rather than of introducing major innovations.

It is quite beyond the scope of this book to summarize the procedures that are followed in the interpretation of the record of a person's responses to the blots. The method of interpretation is extremely difficult and requires an expert knowledge.

Reliability and validity of Rorschach technique are difficult to investigate because Rorschachites insist that no category should be considered in isolation. If it were possible to interpret single scores meaningfully, without reference to other scores, the problem of validation would be a relatively simple one. It would involve the comparison of scores such as number of movement responses with the type of behaviour which the score is supposed to predict. However, this simple situation does not exist, and validation must consist of demonstrating the relationship between the total behaviour pattern manifested in the test situation and the pattern in other situation. ✓

The skilled Rorschach tester can certainly help in the differential diagnosis of mental patients. He seems, however, to be less able to make valid predictions about normal persons.

There are some more projective techniques which although are not much used yet gives sufficient evidence regarding the persons adjustment. A brief mention of those has been made below :—

(*a*) *Self-Description*—Here the subject is asked to write two short descriptions of himself, first as seen by a good friend, secondly by a candid critic. These are interpreted by the psychologist or psychiatrist in conjunction with material gathered during the interview and with other projection test responses.

(*b*) *Literary Productions*—It is a truism that the creative writer, painter, musician, scientist or philosopher expresses his personality in the style and content of his productions. The psychologists interpret these productions to reveal the individuals personality but these interpretations are subjective.

(c) *Sense of Humour Tests*—The mechanism of projection certainly enters into our appreciation of the comic, and there have been some attempts to base personality tests on the types of jokes appreciated. Eysenck found that when people rank jokes or other humorous material in order of funniness, or rate their appreciation, there is remarkably little agreement between them. Some persons, however, consistently tend to rate the simpler sexual and aggressive jokes more highly, whereas other consistently prefer more subtle and clever humour.

In general, the projective tests should not be used by the ameteur. They must be administered with great care. Scoring is difficult and exacting, and interpretation is exceedingly complex.

The constancy of personality

Do personality or social aptitudes tend to remain with us relatively unchanged or do they fluctuate with the passing years ? This is, of course, a fundamental question. If personality fluctuates widely from day-to-day, from situation to situation, then there is no great point in studying it. Instead of studying personality we should merely study the situation on which personal reaction depend.

From our general impressions, most of us assume that personalities do not fluctuate widely. From our ordinary observations, we feel we can depend on people to keep their present personalities for some time. Within limits, we can predict the future reactions of different people in different situations. We say, "Kishore would have liked this theme." Or "Hari would have appreciated this scene." Similarly, we can imagine how another person would be shocked, or baffled, or bewildered by another situation. True enough, we are prepared to see our predictions go unfulfilled upon occasion. The shy lad may react to the 'encouraging debate' in a more adequate manner than one anticipated.

There are certain studies which tend to confirm that personalities do not fluctuate widely.

An interesting study of this type at the nursery school level is reported by McKinnon[16]. In this study sixteen nursery-school children were observed and the predominant characteristic of each child determined. Four main characteristics were used, namely, conformity, caution, aggression, and withdrawal. Some five or six years later the same group was reclassified. Ten of the sixteen children were still in their original group. These results suggest a fair degree of consistency, even at the nursery school level.

There is another investigation in which six elderly adults of the same family could be matched with the personality estimates derived from a diary kept by the mother when they were young children[17].

From this discussion it is clear that personality is a quality of the individual and is not merely the result of the momentary situation in which he finds himself.

---

[16] McKinnon, Kathern M. : *Consistency and Change in Behaviour Manifestations*, Child Development Monograph, 1942, No. 30.

[17] Smith, Madorah, E.A. : *A Comparison of Certain Personality traits as rated in the same individuals in childhood and fifty years later* ; Child Development 1952, 23, pp. 159-180.

CHAPTER 11

ANALYZING AND IMPROVING TESTS

Analysis of Test Items

THAT TEST IS SAID TO BE USEFUL IN WHICH EACH ITEM discriminates or distinguishes between the bright and poor students. The good test should also sample the objectives being measured.

An item analysis is helpful to the teachers or the constructors of the tests for the following reasons :

(a) The poor items can be discarded without imparing the value of the test. Thus much time and labour is saved.

(b) It is useful in diagnosing the nature of the individual difficulties of the students. Remedial instructions can be given accordingly.

(c) Some discarded items of the former tests can be used in the latter tests where they prove to be useful.

(d) The instructor can easily know which items are good for being put in the test and which to be discarded.

For determining the value of a particular item, statistical methods are used but for ordinary purposes we can use the approximation methods.

For analyzing the items we generally divide the test scores into two groups and then compare the individual item responses of each group. The following are the steps used for analyzing the items :—

(1) Determine the criterion of comparison.

(2) Divide the paper into comparison groups.

(3) Tally the individual responses to each item for each group.

(4) Convert each figure to its equivalent percentage.

(5) Compare and interpret the group responses.

(6) Record the results on an individual item card.

Determine the Criterion of Comparison

There are different Criteria of Comparison. Generally the total test score is used as a Criterion for grouping the papers. The students who secures good marks are taken as best students and those securing low marks are taken as poor. Sometimes the test may have two or three parts which are not related and it may happen that one student who secures good marks in one part may secure low marks in the second part and thus his total score on the test will be low. In such instances the grouping would be done on the basis of the scores made on each part (sub-test).

We can also group the test papers of the students on the basis of their scores on previous examinations, their progress reports etc. Students securing better marks in previous examinations should be grouped together and their test papers should be compared with those securing low marks in previous examinations.

Divide the Paper into Comparison Groups

The test papers can be divided in many ways for being put into comparison groups. Some times we divide them in two groups and sometimes in three groups. Most of the test instructors and teachers select the upper and lower quarters for comparison. The middle 50 per cent are disregarded. A few select the upper and lower halves for comparison.

Comparison may also be done by dividing the papers in terms of equal thirds or upper and lower quarters and middle 50 per cent. Some teachers select the top 25 and the bottom 25 papers in large classes, and others select 100 papers, representing the full range of scores.

Tally the Individual Responses to each item for each Group

A tally sheet is prepared which shows the number of examinees who get each item right or wrong or omit it. If the test has two response items say true and false the tally sheet can be prepared as shown below :—

	High-Group—25%			Middle-Group—50%			Low Group—25%		
Item	Right	Wrong	Omit	Right	Wrong	Omit	Right	Wrong	Omit
1	ǀǀǀǀǀǀǀǀǀǀ	////	/	ǀǀǀǀǀǀǀǀǀǀ ///// //	///// ///// /	///	//	//// /////	///
2	///// (5)	//////// (8)	// (2)	//// (4)	////// //////// //////// (22)	//// (4)	/ (1)	///// ///// (10)	//// (4)
3									
4									

The item of each individual is inspected and a tally is put in the column 'Right' if he has responsed correctly, a tally is put in the column 'Wrong' if he has responsed wrongly and a tally is put in the column 'Omit' if he has omitted it. The total of the tallies in each cell may be put within the brackets as shown in the second row of the above tally sheet. In the case of multiple choice items the tally sheet will include the columns for different alternatives and omit, instead of right and wrong or omit. Suppose there are five alternatives, the tally sheet will be as shown below.

Item No.	High Group—Choices						Low Group—Choices						
	A	B	C	D	E	Omit	A	B	C	D	E	Omit	
1	/	///	|||||	||||| |||||	/	//	///	|||||	//// |||||	|||||	/	0	
2	///// ///// (10)	//// (4)	// (2)	/// (3)	/// (3)	0 (0)	//// //// (8)	||||| // (7)	/// (3)	/ (1)	// (2)	/ (1)	
3													

Here A, B, C, D, E are the alternatives for each item. The tally is put in the appropriate column after inspecting each item of different individuals. The column (Omit) records those people who omit the different items. In the above sheet, the middle group can be added. The total of the tallies in each cell is put within the brackets as shown in the second row of the above tally sheet.

Convert each figure to its equivalent percentage

To convert each figure to its equivalent percentage we multiply each figure by $\frac{100}{N}$, where N represents the total number of papers. As in the above table there are 22 papers. To convert 10 to its equivalent percentage we multiply it by $\frac{100}{22}$. Thus its equivalent percentage is 45·4 or 45. The equivalent percentages of other figures of the above table are shown in the next table.

Item No.	High Group—%						Low Group—%						
	A	B	C	D	E	Omit	A	B	C	D	E	Omit	
1	5	14	23	45	5	9	14	23	36	23	5	0	
2	45	18	9	14	14	0	36	32	14	5	9	5	
3													

One caution should be taken while using the percentages, especially in connection with test scores for small classes. The percentages are apt to be misleading unless the raw numbers are kept in mind. For example, if there are only ten papers in a group the percentage will advance by tens. The difference between 50 and 60 percent, then would be the difference of only one paper. This would not be as significant as one might tend to believe if he considers the percentage scores only.

Compare and interpret the group responses

In comparing and interpreting the group responses we try to note how does each item discriminate positively and what is the relative level of difficulty.

If a good number of high group gets an item right which a little number of low group gets right we say that item discriminates positively. If a little number of high group gets an item right which a good number of low group gets right we say that item discriminates negatively. When almost equal number of high group and low group gets an item right we say that item does not discriminate. The following illustration will make it clear :

Item	High Group						Low Group					
	A	B	C	D	E	Omit	A	B	C	D	E	Omit
(1)	2	1	[10]	5	1	1	5	2	[2]	1	6	4
(2)	8	[2]	6	3	0	1	1	[12]	5	1	0	1
(3)	2	4	2	[8]	4	0	2	1	6	[7]	3	1

In the above table we see the alternative [C] is right for the first item. Ten students or 50% of high group got it right where as only two students or 10% of the low group got it right and hence there is a positive discrimination.

In the second item the alternative [B] is right. Two students or 10% of the high group got it correct where as twelve students or 60% of the low group got it correct and hence there is a negative discrimination.

In the case of third item the alternative [D] is right. Eight students or 40% of the high group got it right and seven students or 35% of the low group got it right and hence there is no discrimination.

Now the question arises as how great the difference must be between the upper and lower groups before the extent of discrimination can be said to be significant. In other words, how much difference should there be between the high group and low group in order to have a good item. The following method is used to determine whether a difference is significant between the upper and lower groups on a single item. This is done by finding the percentage difference between the two groups and dividing by the standard error of the difference. This computation is done for each item.

Let N_1 = Number in upper group marking item correctly

X_1 = Total number in upper group.

P_1 = Percentage of upper group marking item correctly.

Q_1 = Percentage of upper group marking item incorrectly

N_2, X_2, P_2, Q_2 = Same designations for lower group.

The procedure is illustrated by using the results for the first item of the above table.

Firstly we find the percentage figures for both groups.

$$P_1 = \frac{10}{20} \times 100 = 50 \qquad Q_1 = 100 - P_1 = 50$$

$$P_2 = \frac{2}{20} \times 100 = 10 \qquad Q_2 = 100 - P_2 = 90$$

Standard error of the difference is—

$$SD_{\text{diff.}} = \sqrt{\frac{P_1Q_1}{X_1} + \frac{P_2Q_2}{X_2}}$$

$$= \sqrt{\frac{50 \times 50}{20} + \frac{10 \times 90}{20}}$$

$$= 13{\cdot}03.$$

Now the percentage difference between the two groups is divided by the standard error of the difference.

$$\frac{P_1 - P_2}{SD_{\text{diff.}}} = \frac{50 - 10}{23{\cdot}03} = \frac{40}{13{\cdot}03} = 3{\cdot}07$$

If the resultant ratio exceeds 2, the item is said to be discriminating. All items having a ratio less than 2 should be revised or eliminated from the test. In the above case the resultant ratio exceeds 2 and hence we can say that the first item is discriminating. Let us consider the third item of above table.

$$P_1 = 40 \qquad Q_1 = 60$$

$$P_2 = 35 \qquad Q_2 = 65$$

$$SD_{\text{diff.}} = \sqrt{\frac{40 \times 60}{20} + \frac{35 \times 65}{20}} = 15{\cdot}2 \text{ (approximately)}$$

$$\therefore \ \frac{P_1 - P_2}{SD_{\text{diff.}}} = \frac{40 - 35}{15{\cdot}2} = \frac{5}{15{\cdot}2} = {\cdot}32$$

This resultant ratio is less than 2 and hence we can say the third item is not discriminating.

These figures might be put into the table form for ready reference. The nature of such a table can be illustrated by using figures from the above examples.

Item No.	Upper Half		Lower Half		$P_1 - P_2$	$SD_{\text{diff.}}$	$\frac{P_1 - P_2}{SD_{\text{diff.}}}$
	N_1	P_1	N_2	P_2			
[1]	10	50	2	10	40	13·03	3·07
[2]							
[3]	8	40	7	35	5	15·2	·32

For approximation we can form the following table to see whether the item is discriminating positively, negatively or not discriminating. Let us illustrate it with the help of above examples.

	Item	% Responding Correctly		
		Upper Group 25%		Lower Group 25%
Positive discrimination.	1	[50]		[10]
Negative discrimination	2	[10]		[60]
No discrimination	3	[40]		[35]

Most of the examples thus far have been based on only the upper and lower groups. The suggestion still holds that it will usually be wise to include also the middle group in making the comparison and interpretations. The procedure need be no different except to add the middle group to the chart.

Let us illustrate the interpretation further with the help of the following chart :

Choice	Upper 25%						Middle 25%						Lower 25%					
	A	B	C	D	E	Omit							A	B	C	D	E	Omit
1	0	0	[90]	0	10	0							15	5	[45]	10	20	5
2	0	20	10	0	[70]	0							0	10	10	0	[60]	20
3	20	50	[15]	15	0	0							10	60	[13]	5	2	10
4	[36]	26	24	0	14	0							[12]	12	40	5	31	0

In the above chart the right responses have been put with in the brackets.

Interpretation—First item

1. The item is discriminating positively because 90 per cent of the upper group responded it rightly and only 45 per cent of the lower group responded it rightly.
2. The item is not helpful in discriminating within the upper group. Almost the whole group selected the correct answer.
3. It is not a difficult item because almost half the total number of lower group responded it correctly, and only 5 per cent of the lower group omitted it. The omissions may be due to lack of knowledge and not due to ambiguity of the item.

4. The item is suitable for being put in this test.

Interpretation—Second item

1. The item is discriminating positively but it is easy. More than half of the upper group as well as lower group selected the correct response.
2. None from the upper or lower group selected the alternatives A and D. They are dead weight in the item. Either these alternatives be left out or modified.

Interpretation—Third item

1. About one seventh of each group selected the right answer. So the item is not discriminating.
2. In each group majority of the students selected choice B than the correct choice C. This indicates that the students who selected the alternative B did not have the correct information.
3. Choice E is not a good detractor. It should be modified.
4. Investigations should be made to see why students selected the choice B. This seems that teaching was faulty. The item needs the modification and is not to be discarded.
5. The item may not appear outwardly to be discriminating. It may, in reality, be a perfect measure of discrimination between those who know the point in question and those who do not.

Interpretation—Fourth item

1. The item is discriminating positively.
2. The item is well discriminating within the upper group.
3. It is relatively difficult because less than half of the upper group responded it correctly.
4. The discrimination is not a perfect one because 12 per cent of the lower group responded it correctly and they might have guessed the right answer.

These four illustrations are not the complete help in item analysis, but they can be used with an advantage while analysing the items.

The test makers should give attention to the objectives which the item purports to measure.

1. What objective is this item supposed to measure?
2. Do the results indicate that it is being measured adequately?

While these aspects should be given careful consideration during the construction of the test, it should also receive attention during the analysis and improvement phase.

Record the item analysis rasults on an individual item card

In order to have a record of the discriminating power and relative difficulty of each item, the item analysis information should be placed on an individual item card. The record can be maintained in the following way.

Item. which one of the following sets of words is associated primarily with foundry work?

(A) expansion and extension.
(B) cope and drag.
(C) tap and die.
(D) flange and fuse.
(E) hollow mandrel and blow horn.

	High Group [20 in this group]						Low Group [20 in this Group]					
Choice	A	B	C	D	E	Omit	A	B	C	D	E	Omit
No. right	2	1	[10]	5	1	1	5	2	[2]	1	6	4
% right	10	5	[50]	25	5	5	25	10	(10)	5	30	20

CHAPTER 12

CUMULATIVE RECORD CARDS

A place for cumulative record cards

IN THE FIRST CHAPTER WE STARTED BY DEFINING THE AIM OF education as the development of a well-integrated person, capable of exercising such responsibility in society as his powers allow. It was then pointed out that some sort of evaluation instruments are necessary to see whether the individual is developing fully or not. Generally school examinations are used as the evaluation instruments. We have already discussed the limitations of the school examinations in some previous chapter. The relatively short examination which takes, as it were, a snap sample of a pupils' abilities and attainments has other disadvantages.

First, it is so brief when compared with a pupils years of progress in the school that it may miss the changes in his development. Secondly, when used for purposes of selection it takes in the quality of a hurdle which has to be cleared. It may bring about an undesirable build-up in a pupil as a result of his fear of consequences of failure. Thirdly, the impartiability of the large-scale mass examination consisting of standardised tests may be an illusion and turn out to be mere insensitivity to the emotional differences in the children. Underneath the apparent uniformity individuals may be reacting quite diversely with corresponding results in their test performance. Some are confident, others a little shaken, and a few paralysed.

The only way to mitigate these evils of school examination is not to abolish them altogether, but to put them in their proper place and to find the means to take into consideration and give due credit for the work done in the class-room and for progress made by a child throughout the school course. Recognition of these facts and of the need to make the fullest use of the teacher's fairly long standing acquaintance with his pupils has led educationists to suggest that a child's total progress should be recorded by the teacher. This should be done on his 'school record card'. The Secondary Education Commission (India) felt the need of the progress of the pupil during his study period. In the words of the Commission's report[1] "Neither the external examination nor the internal examination, singly or together, can give a correct and complete picture of a pupil's all round progress at any particular stage of his education ; yet it is important for us to assess this, in order to determine his future course of study, or his future vocation. For this purpose a proper system of school records should be maintained for every pupil indicating the work done by him in the school from day to day, month to month, term to term, and year to year. Such a school record will present a clear and continuous statement of the attainments of the child in different intellectual pursuits throughout the successive stages of his education. It will also contain a progressive evaluation of development in other directions of no less importance, such as the growth of his interests, aptitudes, personality traits, his social adjustments, the practical and social activities in which he takes part. In other words it will give a complete career."

The maintaining of pupils records is not a new discovery. Testing and recording had a prominent place in the schools during the past as well. Some teachers are familiar with the term 'permanent record' as a designation for the official record of pupil progress. In many schools though the teachers maintains cumulative records of the pupils yet they call it by another name. In England many heads have long used them for advising pupils on a choice of vocation ; but in the educational reforms following upon the 1944 Education Act it was advocated that records should ultimately supplant single examination for purposes of selection and guidance.

---

[1] Report of Secondary Education Commission, Govt. of India, 1953.

Purpose of Record Cards

Their general object is to provide more useful information about the pupils than is possible under examination conditions. "They are written running commantries, dynamic and cumulative, of the child's years of school life. This material may be used for information only, or for prognosis, or for diagnostic purposes".[2] Wendell C. Allen says, "To be cumulative the basic permanent record of pupil progress and adjustment in school should be a unified, usable outline presenting a well rounded picture of experience and development. Concerned with skills, personality, interests and activities, a cumulative record should be not only a means of studying the individual pupil but a result of such study. Essentially, the cumulative record should be, and should function primarily as a guidance tool rather than merely as an instrument for official record. In any school situation it should be a tool which facilitates coordination of all the personnel activities carried on by the staff and should be of real value to agencies that work with individuals after they have left school. An adequate cumulative pupil record provides the data essential to study of a student's present situation in the light of his past development. It is as well, a record sufficiently rich in detail and presented in such form that the school staff and research specialists may profitably utilize its data for research purposes. Cumulative record data can be useful to research in many areas of education".[3] It is often customary to prepare a record card for each stage of a pupil's schooling—in the infant school, the junior school and the secondary school. The information on all the cards may be used diagnostically, for knowledge of child's strengths and weaknesses guides teaching, but each card has in addition a more specific purpose.

---

[2] Peel, E. A. : *The Psychological Basis of Education,* Oliver and Boyd Edinburgh : Tweeddale Court, London : 39A Welbeck Street, W. 1, 1956.

The infant school record card is intended for information mainly to record the child's adjustment to the group life of school which he has newly joined from the security of family protection. It records also his first responses to number and words. So we get a glimpse of his emotional and social development and of his early intellectual showing. Both these pieces of information may be very valuable in case of later failures to develop normally.

The special purpose of the junior school record card is to provide the information needed for allocating pupils to secondary schools. It has to help in educational guidance.

The principal aim of the secondary school card is to furnish details of abilities, attainment, aptitudes and interests that would enable us to advise the pupil about a vocation. Its main objective is vocational guidance.

Contents of Record Cards

From what has been said above it may be concluded that the contents of record cards should be different at the infant school level, junior school level, and secondary school level.

Infant School. At this level we are mainly concerned in recording the emotional qualities of the infant, viz., cooperation with adults, co-operation with other children, self-confidence, stability of temperament and persistence at school tasks.

At this level the subjects are taught in an informal way, rather we hardly use the word 'subject'. Hence our entries are limited to bodily skills and co-ordination, use of language, clarity of speech, drawing, dancing, singing etc.

---

[3] Allen, W. C.: *Cumulative Pupil Records*, Bureau of Publications, Teachers College, Columbia University, New-York, 1948.

We may, however, record the intellectual powers of the infant. For measuring the intellectual powers we should use group pictorial tests because the infant is not familiar with class regimentation and moreover he lacks in the reading and writing ability.

Junior School. In England junior school record cards are very helpful in allocating pupils to different types of secondary schools, viz., Grammar, Secondary Modern, and Technical Schools. Pupils are sent to these schools according to their age, ability and aptitude. In India there is no system for allocating pupils to the secondary schools according to their ability, aptitude, interest and attainment in school subjects. No doubt educational practices are not portable commodities yet no one shall deny that there are educational practices which can be used universally. India can take a hint from England for the allocation of junior school pupils for various secondary school courses.

The junior school record card contains entries about the child's intelligence and attainments. His interests are also recorded. Interest may be intellectual, practical, aesthetic, social and physical. For our purpose we may take into account intellectual or academic and practical interest.

As we may expect, the temperamental qualities we look for are not quite the same as those recorded in the infant school record card. Apart from a note on the prevailing attitude of the child, absorbed in self or highly interested in the world, these are six qualities which may be listed on the record card in three pairs. The first pair, self-confidence and self-criticism, concern an attitude to 'self'. The second pair, sociability and co-operation, present an attitude to 'others', and the third pair, perseverance and conscientiousness, an attitude towards work.

A column for home conditions and physical health of the child is also used. Those columns are helpful for border line cases.

Secondary Schools. In the secondary schools many subjects are taught and as such the entries relating to school attainments may be elaborated and presented under several headings like History, Geography, Language, Physics, Chemistry, Algebra, Geometry, Arts, Agriculture etc. These entries are also helpful for detecting any special aptitudes which the pupil may possess. More space is given to pupil's preferences, aims and interests. There can be introduced a section in the record cards which may be helpful for vocational guidance.

Most pupils take up a job on completing their secondary education. In England the routes of employment are various—many pupils find jobs without the help of head teachers, many are helped directly by the heads and many need to seek the advice of Youth Employment Officers. In India Employment Officers are doing their best to direct pupils for various jobs. Employment Officers can do their job best if the relevant information about the pupils may be passed on to them by the heads of the institution.

Examples of the vocational guidance section in the Record Card

(1) Athletics

(2) School offices held

(3) Leisure interests :

    (A) Intellectual

    (B) Practical

    (C) Social

(4) Special aptitudes

(5) Family occupations :

    (A) Father

    (B) Brothers/Sisters

(6) Employment suggested by :

    (A) Parents

    (B) Pupil

(7) Unsuitable employment

Physique

General appearance		

Vocational Tests

Type	Result

(8) Type of employment recommended

(9) Subsequent history

To sum up the school record should be able to facilitate guidance, whether it be educational or vocational ; it should help to discover the personality and behaviour problems of children with their underlying causes ; it may be easily used in planning the curriculum in terms of the needs ; interests and abilities of children. The maintaining of school records will be possible if the parents co-operate. For this establishment of teacher-parents association is very essential.

In the end of this chapter a sample of the school record has been given. Many of the entries that are used in the record cards in England are missing in this record card. Here we have no entry for intellectual progress of the pupil. The reason for the omission of this entry is the dearth of intelligence tests in this country. Moreover, most of the school teachers are not familiar with the administration and interpretation of these tests. Therefore, the assessment of I.Q.s have been left out in the record cards in India. The class teacher can assess the rough levels of intelligence in order to classify them into three categories: the bright, the average or normal, and the backward or subnormal ; and a teacher who knows his children well enough will be able to do so without much difficulty.

The filling in of the record card is a difficult job. It requires the keen observation and a perfect training on the part of teacher. It is always better to leave any item blank than to fill in a wrong thing. Sometimes the teachers can consult one another to fill in a certain item. All entries should have the initials of the teacher and the date.

Instructions for Entries—*General Data.*

These entries can be made with the help of school office registers.

1. **Attendance.** This can be noted from the attendance register. The word 'possible' means the number of days the class worked.

2. Position of Responsibility Whether a Monitor, Captain of a team, Secretary of Literary Club etc.

3. Socio economic Background. The teacher can gather this information from the guardians of the pupils. Whether pupil is :father-less, or mother-less, poor or rich or his relations are in high official places etc.

4. Physico medical Report. The height, weight and chest measurements etc. should be entered. Teacher should also note whether the pupil was ill for a certain period during his study course. Nature of illness should :be mentioned. Any physical disability as left handedness, short sightedness, hard hearing etc. should be rated as good, indifferent and poor.

5. Notable Interests. Whether the boy is interested in games, scouting, community work, debate and swimming etc. is to be filled in. Interests can be divided in many categories as shown in the record card. If the boy is interested in writing a poem it means his interest is intellectual.

6. Aspects of General Ability. The six important aspects of general ability are shown in the sample record card. We can rank the boy for each ability as 'Marked', 'Average' and 'Below average'.

7. Special Disabilities. Examples of special disabilities are shyness, inferiority complex etc.

8. Special Skills. Examples of special skills are good singer, excellent player etc. A mention of :any special skill of the pupil should be made in the record card.

9. Personality Traits. The various categories have been shown in the record card. For ordinary purposes the strength may be measured under three categories : (1) strong or marked (2) normal, or average, (3) weak.

10. Pupil's Attitude to School and Preferences. Here we are to see whether the pupil is co-operative or not, takes interest in school functions or not, volunteers for any cause of the school or not, etc.

11. School Work Against each group of subjects fill in the names of specific subjects, e.g., against Social studies, History, Geography, Civics etc.

The assessment will be in two forms; an estimate of the pupil's position on a five point scale, and the marks obtained by him in the examinations. If terminal examinations are held then the average marks of all the terminal and final examinations should be entered.

12. Principal's (or Headmaster's) Remarks. Here the headmaster needs to give his own judgment regarding the school career of the pupil. If necessary the headmaster can take the help of the teachers report regarding the pupil.

Confidential.

SCHOOL RECORD[4]

General Data

Name in full (in block letters, surname first)..........................

Date of birth in figures

Day	Month	Year

Father's name.............................

Guardian's name.....................................Relationship with

pupil..

Address..

....

....

,, Changed to..........................Date....................

,, ,, ,,.......................... ,, ...

,, ,, ,,.......................... ,, ...

,, ,, ,,..

....

School....................................Date of entry....................

Admission Register No..............Date of leaving....................

Reasons for transfer..

To School..........................Date of entry....................

[4] A. N. Basu (Central Institute of Education, Delhi); *School Record* (with a sample record form and instructions of filling it).

1. Attendance

Year	Possible	Actual	Nature of long absence, if any, with reasons
19			
19			
19			

2. Position of Responsibility Held in School, Special Merit Obtained, Etc.

19	
19	
19	

3. Socio-Economic Background

Special Home Conditions	Pupil's position in the family	
Occupation of Parents Father— Mother— Economic Status Comments	Boy 1 2 3 4 6 7 8	Girl

19

19

19

4. Physico-Medical Report

Past History—

Year & Month	Height	Weight	Normal	Contracted	Expanded	Any serious illness	Any special physical disability	General health rating
19 Jan. July								
19 Jan. July								
19 Jan. July								

5. Notable Interests

Categories	19	19	19
1. Physical			
2. Intellectual			
3. Practical			
4. Artistic			
5. Social			
6. Recreational			

6 Aspects of General Ability

Categories	19	19	19
1. Verbal facility			
2. Reasoning			
3. Speed of work			
4. Observation			
5. Practical			
6. Artistic			

7. Special Disabilities

19	19	19

8. Special Skills

19	19	19

9. Personality Traits

Nature	19	19	19
1. General attitude			
2. Self-confidence			
3. Perseverance			
4. Co-operativeness			
5. Sociability			
6. Courtesy & Consideration for others			
7. Emotional Control and stability			
8. Sense of responsibility			
9. Integrity			
10. Initiative			
11. Habit of work			
12. Leadership			

10. Pupils Attitude to School and Preferences

Point of enquiry	19	19	19
Pupil's general attitude to school			
Favourite school subject (s) and schooling			
Subject (s) liked least			
Present educational preferences			
Present Vocational Preference			
Parent's view on child's future			
Teacher's Comment			

11. School Work

Assessment under this section will consists of a general estimate and marks obtained in the examinations. For giving estimate of the pupil's place in the class arrange pupils into five categories : A [well above average], B [above average], C [average], D [Below average], and E [well below average]. Enter the letter of the group within which the pupil falls in subjects concerned ; under the column 'marks' give the average scored by the pupil in the annual and other periodic examinations.

Group	Subject	19		19		19	
		Exam.	Marks	Exam.	Marks	Exam.	Marks
Language & literature							
Social studies							
Sciences							
Mathematics							
Craft							
Art							
Music							
Physical Education							
Other subjects							

12. Principal's (or Headmaster's) Remarks

19	
19	
19	

CHAPTER 13

INTRODUCTION

Origin of Statistics. THE WORD 'STATISTICS' IS BARELY A century old in its present meaning. It seems to have been derived either from Latin **'Status'** or the Italian **'Statista'**, both meaning **'Political State'**. The definition of the word 'Statistics' is given in 'The Elements of Universal Erudition' by Baron J. F. Von Bielfeld as "The Science that teaches us what is the political arrangement of all the modern States of the known world". In olden days kings and ruling chiefs used to take census of population and wealth within their domain to calculate their manpower and strength. They used to collect the data of income, death rates, and birth rates etc. of their country which helped them for better administration. Naturally the compilation of data of facts and figures gave birth to 'Statistics'. History reveals that, during the Maurya Period in India the work of registration of death rates and birth rates was started. Again during the *Muslim Period* they carried on the Statistical surveys. It is therefore, evident that statistics was originated in India long back.

Meaning of Statistics. By 'statistics' we mean quantitative data affected to a marked extent by a multiplicity of causes. By 'Statistical methods' we mean Methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes.

By theory of statistics or more briefly, by 'statistics' we mean the exposition of statistical methods.

W. G. Sutcliffe in his 'Elementary Statistical Methods' says, "Statistics comprises the collection, tabulation, presentation and analysis of an aggregate of facts, collected in a methodical manner, without bias and related to a predetermined purpose."

Development of Statistics. The ancient astronomers used to collect the astronomical data. They studied and collected the record of the motion of planets and heavenly bodies. Naturally statistics began to develop. During the 17th century with the origin of life insurance scheme, statistics came in much prominence. During 18th century J. P. Sussmilch, Prussian clergyman, tried to demonstrate the doctrine of 'Natural order' statistically, in an important Publication John Grant Petty and Sussmilch contributed much towards statistics during the 17th and 18th century under the name of 'Political Arithmetic', which worked as a right arm for the then Government. Soon after the game of chance originated the theory of probability and statistics played its full role for the development of this pronounced theory of Mathematics. It was discovered during the time of Quelet that deviation of different observations from the average of the observations follow the 'Binomial Law' and upon this factor depends the 'theory of statistics'. In the later years people began to collect the data of population, consumption, death rates, birth rates, production, which helped them to have a clear view of Economic conditions of the country on a piece of paper. Thus statistics appeared in the field of Economics too. So much so that the Statisticians were then regarded as Economists. But the theory of Statistics was based on Mathematics and so some regarded Statisticians as Mathematicians. To-day the scope of Statistics is not only upto Economics and Mathematics, but it has extended its scope to Education, Physics, Biology, Astronomy, Anthropology, Meterology, Agriculture and medicine etc.

Methods of Statistics

In statistics we come across to a large number of figures and it becomes very difficult for the brain to comprehend the data easily. For this few devices have been originated with the help of which the mind can grasp the significance of the large number of data. These devices are called 'Methods of Statistics'.

(1) The first step is to collect the data. If the population from which the data is to be collected is very large, then we take a sample and collect the data from this representative sample.

(2) After the collection of data it is classified. Classification condenses our data.

(3) Then we tabulate this classified data. Tabulation gives such a shape to our data which the mind can easily comprehend.

(4) If we like to make the data more comprehensible then we use diagrams and graphs. This is the best way of representing the long data in such a manner so that the significance of it may be grasped by the mind easily.

(5) When different series of data are given to us we try to compare the behaviour of one series to another. Say, if we know the average age of an Indian as calculated from the data collected in India, we may try to compare this average age with the average age of an American. Then we draw out the conclusions on what factors this behaviour varies. Say, as in above case our conclusion may be that the average age of an Indian is less than the average age of an American. Then we may conclude that age depends upon nourishment and nourishment depends upon economic conditions. Naturally we can say economic conditions of India are poorer than that of America.

(6) Statistical methods, sometimes help us in forecasting the conditions which may occur. If we have a record of death rate of population and if we know in later years that death rate is increasing we can predict that in few years the population will become very thin.

The Classroom Teacher and Scientific Movement in Education

The scientific movement in education is an important step in the improvement of the educational system. Many scientific procedures were introduced in educational system. The teacher is to play a definite role in the development of this programme of scientific education. The class-teacher is entrusted with the work of research in the class-room. The research may be connected with class-room, with the students and with the school. He is to study the character, the ability, the health, the interest of the pupils and care for their proper development.

Class Teacher and Statistics

The class teacher cannot fulfill his duty as discussed under this scheme of Scientific Movement in Education unless he understands statistics. The teacher should have a knowledge of statistical methods. Ezekiel has written a statement which concerns the relation between statistics and the class-room research.

He says, "The place of statistical analysis in scientific research is not different from the place of any other technical aid the investigator may apply. It furnishes a means of specially measuring the elements that may be involved and of examining the way they are related ; but it does not of itself furnish an explanation of phenomena, except in so far as the effort to reduce the variables to specific numerical statement, definitely related forces the investigator to think more clearly and definitely about his problem, statistical analysis is not a substitute for logical analysis, clear thinking, and full knowledge of a problem. The methods of analyzing complicated relations.........furnish.........a keen tool or investigating complicated relationships but like all keen tools, they may yield unsatisfactory results or misleading results if employed carelessly or headlessly. Statistical analysis is not a substitute for careful thinking and skilled workmanship in research work ; instead, it is an aid which may make that thought and skill even more productive of worth-while results."

Hardy says, "A philosophy of research, inductively derived, will serve also to prevent the wrong interpretation of that data of research. Sometimes the right thing to do is exactly the opposite of that which the findings of research would seem to dictate. Alfred Adler makes this clear in citing the case of Beethoven, whose deafness would have led a certain type of researcher, had there been such in Beethoven's day, to councel him away from music. Perhaps the best way to put it is that research is not safe in the hands of the more technician in graphics, statistics and mathematics, that it is safe only when guided and used by those who, while masters of technique, are also master interpreters of significance, the meanings, of the products of their work—who are, in short, philosophers because they strive to get all the facts in their meanings."

Here we are not pressing the point that a teacher should necessarily be a researcher and should have a complete knowledge of statistics and statistical methods. But we are definitely making a point that a teacher should have some working knowledge of statistical procedures and should utilize them in assessing or evaluating the progress of the pupils and in comparing their achievements in different school subjects or in different abilities.

There are numerous occasions when a teacher wants to compare the achievement of the pupils of his class in a particular subject with the achievement of the pupils of the same class either in different sections of the same school or in the different schools. This comparison is not possible unless the teacher has the knowledge of preparing a frequency distribution and unless he is familiar with the techniques of calculating the mean, the standard deviation, and the significance of the difference between the two measures.

The teacher uses the different kinds of educational and psychological tests for the evaluation of the development of the pupils. The construction of these tests and the interpretations of the scores obtained by pupils in these tests demands a good working knowledge of statistics on the part of the teacher. For this he should be familiar with the different types of distribution, different kinds of central tendencies and variabilities and different kinds of correlation techniques. He should be familiar with the different kinds of scores like the 'raw score', the 'standard scores' the 'percentile scores' etc. He should have a knowledge of ranking techniques. He should have a knowledge of sampling theory.

At times the teacher may like to do some sort of action research in his class. For example he may like to compare the two methods of teaching a subject. For this purpose he should know how to equate the two groups on the basis of certain essential abilities and how to control the different variables. He should have the knowledge of a normal distribution which may form the basis for making equivalent groups.

Examples of this kind can be multiplied and the importance of statistical methods for teacher can be examplified endlessly. But the purpose of this book is not to write a thesis on the importance of statistics for a teacher. Perhaps after mastering some of the statistical techniques mentioned in this book the teacher may himself realise the importance of these techniques for his profession and this is where the author wants to lead his readers to.

In the end one thing must be said and it is that statistical methods are most dangerous tools in the hand of inexperts. A teacher who uses these methods with limited knowledge may do more harm than good to the educational development of his pupils. Here we should keep in mind a word of caution that one should not come to a definite conclusion simply because his statistical results reveal that fact. Statistical results should always be used with certain precautions.

CHAPTER 14

THE FREQUENCY DISTRIBUTION

Attributes and Variables

We generally talk of two kinds of statistics : the statistics of attributes' and 'the statistics of variables.' An observer may note down the presence or absence of certain attribute in a man or in a series of objects. Say he may see how many persons are '*blind*' and how many '*not blind*' in a given population. Thus we see a few persons possess one attribute while the rest do not. Similarly we may find in a group of persons, if how many are '*Hindus*' and how many '*Non Hindus*'. This sort of methods in statistics are called 'statistics of attribute'.

On the other hand the observer may measure the magnitude of variable character in different individuals or different objects. Say he may find out the ages of different persons at their deaths or the I. Q. of different children at the age of ten. The methods applicable to this kind of observations are called 'statistics of variables.'

Discrete and Continuous Variables

Variables are those measurable quantities which vary from one another. Variables are divided into two classes – Discrete and Continuous.

A continuous variable can attain any value within a certain range. Examples of such variables are birth rate, death-rate, temperature at different places, I. Q. of persons, etc.

A discrete variable is that which attains only discrete values. Example of it is 'petals on flowers', 'family size' etc.

The Process of Classification

If we have a large number of scores or the data of tests we cannot judge its significance only by looking at it. Suppose we have scores of 90 examinees in English, we cannot at once say as to how many of them fall in the best category (say who scored more than 60 marks) ; how many fall in normal category (say who scored more than 30 marks and less than 60 marks) ; how many of them fall in the worst category (say who scored less than 30 marks). But if we divide them in 3 groups, first consisting of examinees getting marks between 60-89, second of those getting between 30-59 and the third of those getting marks between 0-29 ; we can at once judge the number of students falling in each group. This is the simplest idea of classification. This type of grouping mainly consists of these steps.

(1) First we find out the 'Range'. 'Range' is the difference of the highest and the lowest score in the given data. Say if in the above case the highest marks obtained by one examinee is 80 and lowest marks obtained by another examinee is 10 then range will be $(80-10)=70$. (See Example 1).

(2) After that we determine the class-interval. This depends upon the range and the number of intervals we want.

Generally we should see that the number of class-intervals should neither be less than 10 and nor more than 20 but this rule does not hold true in cases where scores are very large or very small. As in example below range is 70 and for our facility and by trial we can take the size of class-interval as of 5 units because it gives us 15 class-intervals. This number 15 is one more than the number 14 which we get by dividing the range by the size of the class-interval. Thus in the present case we have 15 class-intervals which are of course in accord to the above rule. Thus it is advisable to choose the size of class-interval such which may give the number of class-intervals between 10 and 20.

(3) After finding out the number of class-intervals we tabulate the scores in their proper intervals. For this we make

three columns. First column is named class-interval, second as tallies and third as frequency.

In the first column we put down the lowest score at the bottom, a dash infront of it and then the figure which we get after adding the size of class-interval to the lowest score minus one. In the present case lowest score is 10 and infront of 10 we put down $(10+4)=14$ where 5 is the size of the class-interval. This means this class-interval comprises of scores 10, 11, 12, 13, 14. Above this we have a class interval 15-19. Again here 15 is the lowest score of this class-interval because scores upto 14 have been taken in lowest interval. To 15 we add four (the size of the class-interval minus one) and get 19. This second class-interval comprises of scores 15, 16, 17, 18, 19. In this way we go on dividing the scores in the class-intervals till at the top we get 80 - 84 which comprises of scores 80, 81, 82, 83, 84.

It is to be noted that the interval 75—79 contains scores 75, 76, 77, 78, 79 and not 80. But we see among our scores 80 is one of them so we must have one class-interval more, i.e., 80—84 which may satisfy our demand. In the second column we mark the 'tallies'. Say our first score is 15 and we see the class-interval 15—19 contains this score. So we put one mark (/) against this class-interval which is called a tally. Second score is 18. This also falls in the same class-interval. Similarly score 65 will fall in the class-interval 65—69. Care should be taken, it will not fall in the interval 60—64. Similarly score 20 will fall in class-interval 20 - 24 and not in 15 - 19. When against any class-interval we get five tallies we put them as /////. If the number of tallies becomes six, we may write them as (//////) and if it increases to ten we write them as (/////) (/////) and so on. We count the tallies against each class interval and call it a 'frequency' of that class-interval. This frequency is put down in third column. Thus in our present example frequency of upper most class-interval is one. Frequency of class-interval 60—64 is two and so on. We add the total number of frequencies and see the total is 90, i.e., which is the number of examinees. Thus this total number of frequencies tell us whether our tabulation is correct or not.

The manner in which the class frequencies are distributed over the class intervals is called frequency distribution. The readers may see the following example carefully for clear understanding.

Example 1

There are 90 examinees whose scores in English are as below. Group these scores in a frequency distribution.

TABLE 1

The original scores ungrouped

15, 18, 17, 16, 14, 20, 22, 10, 50, 45, 60, 65,
70, 75, 78, 80, 14, 18, 16, 15, 35, 30, 25, 46, 45,
48, 44, 32, 33, 34, 56, 57, 62, 65, 71, 72, 73, 36,
37, 38, 25, 23, 24, 21, 19, 16, 14, 25, 28, 30, 29,
29, 29, 40, 40, 40, 45, 36, 37, 37, 37, 39, 39, 41,
41, 32, 32, 32, 32, 35, 35, 53, 66, 66, 66, 39, 39,
28, 29, 30, 42, 45, 46, 22, 25, 29, 36, 35, 39, 42.

The same 90 scores grouped into a frequency distribution.

Class interval	Tallies	Frequency
80—84	/	1
75—79	//	2
70—74	////	4
65—69	(/////)	5
60—64	//	2
55—59	//	2
50—54	//	2
45—49	(/////)//	7
40—44	(/////) ///	8
35—39	(/////) /////)(/////)//	17
30—34	(/////)(/////)	10
25—29	(/////) /////)/	11
20—24	(/////)/	6
15—19	(/////)///	8
10—14	(/////)	5
		Total = 90

LIMIT OF THE CLASS-INTERVAL IN A FREQUENCY DISTRIBUTION

The following are the three ways of writing the class-intervals.

(a) 40—50
30—40
20—30
10—20

It means that class interval 10—20 contains all the scores between 10 and 20 including the score 10 but excluding the score 20.

The same class intervals can be written as below—

(b) 40—49
30—39
20—29
10—19

It means that class-interval 10—19 contains all the scores between 10 and 19 including 10 and 19 as well.

The same class-intervals can be written as—

(c) 39·5—49·5
29·5—39·5
19·5—29·5
9·5—19·5

It must be clearly understood by the readers that the precised lower limit of the lowest interval is 9·5 and not 10. Again its upper limit is 19·5 and not 19.

In all the above cases the length of the class-intervals is 10. *Method* (c) is correct but it requires more time for the calculation and moreover there are chances of errors in such calculations. *Method* (b) is quite easy one because in this case tabulation work becomes easy but we must be aware that the lower limits and the upper limits are not those which we have written down in the class-interval of (b) but those which we have written down in (c). *Method* (a) of writing the class-intervals is accurate but the presence of the same two numbers twice sometimes mislead us in tabulating a score. Say we want to have a tally for the score 20. For this we should put tally against the class interval 20—30 but sometimes due to the appearance of 20 in the interval 10—20 we put the tally against this interval. If we can avoid this confusion then of course this method is quite accurate.

Method for finding the Mid. Point of Class-Interval

Middle point of the class-interval =

$$\text{Lower limit of the class-interval} + \frac{(\text{upper limit} - \text{lower limit})}{2}$$

Say in the above case if we want to find the middle point of the class interval 10—19 we should note, its lower limit is 9·5 and upper limit is 19·5.

$$\therefore \text{ Mid. point of class-interval } (10-19) = 9{\cdot}5 + \frac{(19{\cdot}5 - 9{\cdot}5)}{2} = 9{\cdot}5 + 5 = 14{\cdot}5$$

Now let us have a few more examples of frequency distribution.

Example 2 (a)

Tabulate the teachers salaries given below in a frequency distribution.

Rs. 50, 45, 80, 85, 82, 73, 75, 96, 40, 100,
91, 65, 79, 43, 56, 57, 95, 92, 49, 53.

The salaries tabulated into a frequency distribution where length of the class-interval = 10 units.

TABLE 2 (a)

Class-intervals	Tallies	Frequency
100—109	/	1
90—99	////	4
80—89	///	3
70—79	///	3
60—69	/	1
50—59	////	4
40—49	////	4
Range —(100—40)—60		Total No. = 20

Length of the class-interval = 10 units

$$\therefore \text{ Total No. of class-intervals} = \frac{60}{10} + 1 = 7$$

TABLE No. 2 (b)

The salaries tabulated into a frequency distribution where length of the class-interval = 5

Class-intervals	Tallies	Frequency
100—104	/	1
95—99	//	2
90—94	//	2
85—39	/	1
80—84	//	2
75—79	//	2
70—74	/	1
65—69	/	1
60—64		0
55—59	//	2
50—54	//	2
45—49	//	2
40—44	//	2
Range = (100—40) = 60		Total No. = 20

Length of the class interval = 5 units

$\therefore$ Total No. of class-intervals $= \frac{60}{5} + 1 = 13.$

Example 3.[1]

Scores obtained by 125 pupils (boys and girls) upon an achievement test in general science. Tabulate them into a frequency distribution.

101, 86, 79, 81, 76, 74, 77, 74, 72, 72, 71, 69, 69, 67, 67, 69, 65, 64, 69, 63, 64, 64, 61, 63, 65, 61, 60, 60, 60, 60, 58, 59, 57, 61, 62, 59, 61, 57, 60, 55, 56, 57, 54, 59, 54, 57, 57, 54, 58, 57, 53, 55, 56, 56, 55, 53, 54, 53, 53, 55, 53, 50, 56, 54, 48, 49, 50, 48, 48, 47, 48, 46, 47, 47, 47, 45, 46, 44, 49, 47, 54, 42, 45, 46, 44, 44, 45, 41, 45, 46, 47, 40, 42, 42, 42, 40, 36, 40, 42, 37, 40, 39, 40, 40, 35, 40, 32, 35, 37, 39, 32, 35, 34, 28, 28, 31, 28, 29, 32, 32, 31, 33, 30, 24, 23.

---

[1] This data has been taken from the author's thesis '*A Comparative Study of the Standard of General Science' In certain Secondary Schools in England and in U. P.* For pupils of a chosen age range, M.A. Education thesis, Birmingham England, 1957.

TABLE 3

Class-interval	Frequency
C. I.	*f*
100—109	1
90—99	0
80—89	1
70—79	9
60—69	23
50—59	32
40—49	36
30—39	7
20—29	6
	Total No. (N) = 125

Length of the class-interval = 10.

SUMMARY

(1) This chapter has dealt with the definition of the statistics of attributes and the statistics of variables.

(2) It has been discussed in this chapter as to what are the discrete variables and what are the continuous variables.

(3) The three steps have been discussed by which a given number of scores can be grouped into a frequency distribution. We should note here that it makes no difference whether we put the lowest class-interval at the bottom or at the top.

EXERCISES

1. Tabulate the following high school enrolments in a frequency distribution.

254	363	161	133	101	39	123	60
199	988	66	53	99	51	205	182
237	91	95	117	73	98	155	61
125	88	49	33	56	28	609	321
101	112	105	291	34	14	74	352

2. Tabulate the following 20 scores into two frequency distribution using (1) an interval of 10 and (2) an interval of 15.

35 36 48 49 53 85 102 108 38 83
76 85 56 69 115 96 120 93 30 46

3. Tabulate the following 15 scores into a frequency distributions and see how many scores lie between 17 and 23 (including 17 but excluding 23)

5 10 12 15 17 21 20
19 6 14 25 13 17 21
24

Ans. 6 scores lie between 17 and 23 (including 17 and excluding 23)

Hints. Take a class-interval of 2 units. See 2 frequencies are in the class-interval 17—19; 2 in 19—21 and 2 in 21—23 (In these intervals 17 is included but 23 is not).

CHAPTER 15

AVERAGES AND OTHER MEASURES OF LOCATION

SOMETIMES WHEN WE WANT TO COMPARE THE DEATH RATES of the population in different countries or compare the progress of students in different schools we tabulate the death-rates of various populations in a frequency distribution or tabulate the progress of students in different schools in a frequency distribution. But here our work is not finished. We then find out the average death-rate of various populations or average progress of students of different classes. These averages are termed as '*measures of central tendencies*'. Generally we deal with three types of central tendencies :

(*a*) *The Arithmetic Mean.*

(*b*) *The Median.*

(*c*) *The Mode.*

Particularly we use the word 'average' for 'Arithmetic Mean' but in general the word average denotes the above three measures of central tendencies. In the present book when we use the word average we will mean 'The Arithmetic Mean.'

(A) ARITHMETIC MEAN OR 'AVERAGE'

When Scores are ungrouped. In short Arithmetic Mean is often spoken as Mean. Its value is calculated by adding all the scores or the other measures and then dividing them by the total number of scores or measures. Say if 8 persons' expenses in a family are as Rs. 20, Rs. 30, Rs. 40, Rs. 50, Rs. 60, Rs. 70, Rs. 80 and Rs. 90 respectively and we want to find out the average expenses of that family we add all the above expenses of the individual members and divide it by total number of members, i.e., by eight.

Thus average expenses of family =

$$\text{Rs. } \frac{(20+30+40+50+60+70+80+90)}{8}$$

$$= \text{Rs. } \frac{440}{8} = \text{Rs. } 55$$

Had the expenses of the individuals been,

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$$

then average expenses of the family would have been

$$\frac{x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8}{8}$$

If the expenses of N individuals in a family be,

$$x_1, x_2, x_3, x_4, \ldots\ldots\ldots X_N$$

then average expenses of the family will be

$$\frac{x_1+x_2+x_3+x_4+\ldots\ldots X_N}{N}$$

This in short we can write as $\frac{\Sigma(x)}{N}$ where N is the total number of scores, x is score of the individual and Σ denotes the sum of all the scores. In the above numerical example x has different values as 20, 30, 40, 50, 60, 70, 80, 90. Therefore, $\Sigma(x)$ means sum of all these figures.

Average when the scores are grouped

For calculating the average (mean) two methods are used. One is called the 'long method' and the other is called the 'short method'.

Long Method. Suppose we test 40 students in Hindi. Their scores are tabulated in a frequency distribution as below. Let us find out the average score of the class of the above students.

TABLE 3 (b)

Class-intervals	Frequency (f)
10—19	15
20—29	12
30—39	10
40—49	3
	Total No. =40

In such grouped scores we firstly find out the middle value of each class interval and put them in a column named middle value of (x). Each middle value of different class intervals should be placed against it in the said column as shown below. Say the middle value of the class-interval

$$40-49 \text{ is } 39{\cdot}5+\frac{(49{\cdot}5-39{\cdot}5)}{2}=39{\cdot}5+5=44{\cdot}5$$

This we have discussed in the previous chapter.

Similarly the mid-value of the class-interval

$$10-19 \text{ is } 9{\cdot}5+\frac{(19{\cdot}5-9{\cdot}5)}{2}=9{\cdot}5+5=14{\cdot}5 \text{ and so on.}$$

Then we multiply the middle value (x) of different class-intervals by the frequency (f) of that class-interval. Say the Middle value of the class-interval 40—49 is 44·5 and the frequency is 3. We multiply 44·5 by 3 and put it under column (fx) and against the class-interval to which this belongs. Similarly in (fx) column of the class-interval 10—19 we will put $14{\cdot}5\times 15=217{\cdot}5$ and so on. Add the fx of different class-intervals and call it $\Sigma(fx)$ where Σ means sum as explained above. Divide this $\Sigma(fx)$ by total number of frequencies, i.e., Σf or N. In the present case $N=40$.

Thus $$\frac{\Sigma fx}{N}=\text{Average}$$

Calculation of average by using the long method

[Data from Table 3 (b)]

Class-interval C.I.	Mid-value x	Frequency f	Product of the frequency and mid-value fx
10—19	14·5	15	217·5
20—29	24·5	12	294
30—39	34·5	10	345
40—49	44·5	3	133·5
		$\Sigma f = N = 40$;	$\Sigma(fx) = 990$

Length of the class-interval = 10.

$$\text{Average} = \frac{\Sigma(fx)}{N} = \frac{990}{40} = 24{\cdot}75.$$

In questions where already mid-values of class-intervals are given it saves one step of finding mid-values.

The following is an example of this kind. Find out the average score for the following data.

TABLE 4

Mid-Value x	Frequency f	Product of Frequency and mid-value fx
4	2	8
8	5	40
12	4	48
16	3	48
	$N = 14$	$\Sigma(fx) = 144$

$$\text{Average} = \frac{\Sigma(fx)}{N} = \frac{144}{14} = 10{\cdot}28$$

Calculation of mean for data of Table 2 (b) by using the long method

Class-intervals C.I.	Mid-value x	Frequency f	(Freq.) (mid-value) (fx)
100—104	102	1	102
95—99	97	2	194
90—94	92	2	184
85—89	87	1	87
80—84	82	2	164
75—79	77	2	154
70—74	72	1	72
65—69	67	1	67
60—64	62	0	0
55—59	57	2	114
50—54	52	2	104
45—49	47	2	94
40—44	42	2	84
		$N=20$	$\Sigma(fx)=1420$

Length of the class interval $=5$

$$M=\text{mean}=\frac{\Sigma(fx)}{N}=\frac{1420}{20}=71.$$

Short Method. In the case of Long Method of calculating the mean we have seen the complication of the work. Although we used a very simple example for discussing that method even then multiplication of decimals occurs in. In the cases where frequencies of the class-intervals are great and also mid-points are big numbers, this method becomes more complicated. To avoid this complication Short. Method of calculating the mean is used which is discussed below.

In this method we firstly assume the mean as the middle value of some class interval. There is no hard and fast rule for it but generally the assumed mean is taken at the centre of the distribution or sometimes we take the assumed mean as the middle value of that class-interval whose frequency is maximum. In either case we get the correct result.

Data from Table 3 (b)

Calculation of mean by short method

Class-interval	Mid-value	Frequency	Deviation of mid-value from assumed mean	Product of f and x'
C.I.	x	f	x'	$=fx'$
10—19	14·5	15	—2	—30
20—29	24·5	12	—1	—12
				—42
30—39	34·5	10	0	0
40—49	44·5	3	1	3
				3
		N=40		$\Sigma(fx')$

$=-42+3$

$=-39$

Length of the Class-interval $=10$

Total Number of frequencies $N=40$

Correction in terms of class interval $=c=\dfrac{-39}{40}$

Correction in terms of score $=c\times$ class-interval

$$=c\times i$$

$$=\frac{-39}{40}\times 10=-\frac{39}{4}$$

AM = Assumed mean $=34{\cdot}5$

$$\therefore \text{ Mean}=M=34{\cdot}5+\left(\frac{-39}{4}\right)=34{\cdot}5+(-9{\cdot}75)=34{\cdot}5-9{\cdot}75$$

$$=24{\cdot}75.$$

In the present case we have taken the assumed mean as the mid-value of the class-interval 30—39 which is almost in the centre of the distribution. We could also take the middle value of the class-interval 10—19 as assumed mean because maximum frequency occurs in this interval.

After assuming the mean we find out the middle values of all the class-intervals and put them under column (x).

The next step is to find out the deviation of these middle values from the assumed mean in terms of class-interval and put them under column x'. This is simply the difference of the middle value of the class-interval and the assumed mean divided by the length of the class-interval. Say in the present case mid-value of the lowest class-interval is 44·5 and the assumed mean is 34·5.

Thus $(44{\cdot}5-34{\cdot}5)=10$ score unit $=1$ interval (because here length of the class-interval $=10$).

Again mid-value of class-interval 30—39 is 34·5 and also 34·5 is assumed mean.

$\therefore$ Deviation $x'=34{\cdot}5-34{\cdot}5=0$ score unit $=0$ interval.

Mid-value of class-interval 20—29 is 24·5

$\therefore$ Deviation from assumed mean

$=(24{\cdot}5-34{\cdot}5)=-10$ score unit $=-1$ interval

Mid-value of class-interval 10—19 is 14·5

$\therefore$ Deviation $x'=14{\cdot}5-34{\cdot}5=-20$ score unit.

$=-2$ intervals.

In our solution we are putting x' in terms of intervals and not in terms of score unit because it will save much of our multiplication. After few steps we can convert our answer from unit of intervals to score units which would be clear from the following steps.

We multiply each deviation by the frequency of that class-interval and put it under column (fx').

Thus in the lowest class-interval frequency is 3 and deviation is 1. So $(fx') = 3 \times 1 = 3$.

Above this interval $f = 10$ and $x' = 0$ $\quad \therefore fx' = 0$
In class-interval 20—29 $f = 12$ and $x' = -1$ $\therefore fx' = -12$
In class-interval 10—19 $f = 15$ and $x' = -2$ $\therefore fx' = -30$

The next step is to add +ve valves of (fx') at one place and −ve values at another. We add both these values and call it $\Sigma(fx')$.

Thus here +ve value is 3 only
and −ve values are −30, −12
$\therefore$ sum of minus values $= -42$

$$\left.\begin{array}{ll}\text{Addidg } 3 + (-42) = -39 \\ \text{Thus} \qquad \Sigma fx' = -39\end{array}\right\}\left\{\begin{array}{l}\text{(This is the algebraic sum} \\ \text{where signs are considered}\end{array}\right\}$$

Dividing $\Sigma fx'$ by the total number of frequencies (N) we get the correction. But as discussed above this correction will be in terms of the class-intervals. To change this correction in terms of the score unit we multiply the correction in unit of intervals by the length of the class-interval 'i'. Thus in the present case correction in terms of the class-intervals

$$= c = \frac{\Sigma fx'}{N} = \frac{-39}{40}$$

ci = correction in terms of score units

$$= \frac{-39}{40} \times 10 = \frac{-39}{4} = -9{\cdot}75.$$

The readers can see that if we had taken deviations in terms of score units and multiplied them by the frequency of the respective class-interval we would have got the same correction as −9·75. To show this we take only the three columns f, x' and fx'

f	x'	fx'
15	-20	$15\times(-20)=-300$
12	-10	$12\times(-10)=-120$ -420
10	0	0
3	10	30 $-\quad 30$
N=40		$\Sigma fx'=-420+30=-390$

$$\therefore \quad \frac{\Sigma fx'}{\text{N}}=\frac{-390}{40}=\frac{-39}{4}=-9{\cdot}75$$

From the above calculations it is evident that if we take deviations in terms of the score units in the beginning the solution becomes difficult. To avoid this we adopt the former procedure. When the correction in terms of the score unit is known we add it to the assumed mean (AM).

Thus here AM$=34{\cdot}5$ and ci (correction in score unit)

$$=-9{\cdot}75$$

The formula for mean by short method is :

$$\text{M}=\text{A.M.}+\frac{\Sigma fx'}{\text{N}}\times i$$

where M=Mean

A.M.=Assumed mean

$\frac{\Sigma fx'}{\text{N}}\times i=$ correction in terms of scores.

In the present case M (Mean)$=34{\cdot}5+(-9{\cdot}75)=34{\cdot}5-9{\cdot}75$

$$=24{\cdot}75$$

This is the the same result which we got by using the long method.

Let us now use a few more examples to illustrate how the short method is used for calculating the mean.

(Data from Table 3)

Calculation of the mean of the scores obtained by 125 pupils upon an achievement test in science.

C.I.	f	x'	fx'
100—109	1	5	5
90—99	0	4	0
80—89	1	3	3
70—79	9	2	18
60—69	23	1	23
50—59	32	0	0
40—49	36	1	—36
30—39	17	2	—34
20—29	6	3	—18
	$N = 125$		$\Sigma fx' = -39$

A.M. (Assumed Mean) $= 54{\cdot}5$

$$c \times i \text{ (correction)} = -\frac{39}{125} \times 10 = -3{\cdot}1$$

$$\text{M (Mean)} = 54{\cdot}5 - 3{\cdot}1 = 51{\cdot}4$$

Calculation of the mean of the scores given in table 2 (b).

Class-interval C.I.	Frequency f	Deviation x'	Product of freq. & dev. (fx')
100—104	1	6	6
95—99	2	5	10
90—94	2	4	8
85—89	1	3	3
80—84	2	2	4
75—79	2	1	2
			33
70—74	1	0	0
65—69	1	—1	—1
60—64	0	—2	0
55—59	2	—3	—6
50—54	2	—4	—8
45—49	2	—5	—10
40—44	2	—6	—12
			—37

length of the class-interval $= i = 5$ $N = 20$ $\Sigma(fx') = 33 + (-37)$
$= -4$

A.M. (Assumed Mean) = 72

c (correction in terms of intervals) $= \frac{\Sigma(fx')}{N} = \frac{-4}{20} = -\frac{1}{5}$

$c \times i$ (correction in terms of score) $= \frac{\Sigma(fx')}{N} \times i$

$$= -\frac{1}{5} \times 5 = -1$$

M (mean) $= AM + ci$

$$= 72 + (-1)$$

$$= 72 - 1 = 71$$

The value of the mean by this method is same as by long method. We can very well see in this example as to what extent short method is easier than long method.

THE MEDIAN (Md.)

Median when Scores are Ungrouped

The median in a series of scores is said to be the middle most or central point of that series when scores are arranged in ascending or descending orders of their magnitudes.

Median when Number of Scores is odd

Suppose we have scores 10, 5, 6, 8, 4, 9, 7. We arrange them in ascending orders of magnitude as 4, 5, 6, 7, 8, 9, 10. Here evidently we see 7 is at the middle point, of which 3 scores are on the left and 3 on the right. This is the median. Here we see that the number of scores is odd and in such cases it is easy to find out the median.

When Number of Scores is Even

When the number of scores is even, e.g., 4, 2, 5, 6, 3, 7 then again we arrange them in ascending order of magnitude as 2, 3, 4, 5, 6, 7. Here total number of scores is 6 which is an even quantity. Here either 4 or 5 can be middle points. But there cannot be two middle points. Middle most score might be lying between 4 and 5. We see that :

Score 4 has lower limit 3·5 and upper limit 4·5

Score 5 has upper limit 5·5 and lower limit 4·5

Thus 4·5 is the middle score of 4 and 5 and thus 4·5 is the middle score of the whole series. This is the median.

The mathematical formula for finding out the median is

Median (M*d*) $= \left(\frac{N+1}{2}\right)^{th}$ measure in order of magnitude where N is total number of scores. Thus in the former case there are 7 scores. Median will be at $\left(\frac{7+1}{2}\right)^{th}$ or 4th score which is evidently 7.

In the second case median will be at $\left(\frac{6+1}{2}\right)^{th}$ score or at 3·5th score. Third score is 4, having upper limit 4·5 and fourth score is 5, having lower limit 4·5. Thus 3·5th score will be 4·5 which is the median.

Graphical idea of Media

The graphical idea of median will be clear in later chapter where central tendencies have been discussed graphically. It will be sufficient here to say that median in case of frequency curve is the value of the variable through which if any vertical is drawn it may divide the area of the curve into two equal parts.

Median when Scores are Grouped

The median in grouped scores is said to be the point of the distribution below which and above which lies 50% of the scores. This means above that point lies half the total number of scores and below that point lies half the total number of scores. This definition will be clear from the following solution of finding the median.

TABLE 5

To find out the Median

Class interval C.I.	Frequency *f*
100—104	4
95—99	2
90—94	2
85—89	5
80—84	3
75—79	2
70—74	1
65—69	3
60—64	10
55—59	2
50—54	2
45—49	2
40—44	2
	Total=40

Here we see total number of scores is 40. Thus half of the total number of scores will be 20. Now we start to find out the point below which lies 20 scores and above which lies 20 scores. We go on counting the score from below and while reaching the class-interval 60—64 we see, upto this interval we have 18 scores or frequencies. But we should keep in mind, class-interval 60—64 means, scores 64·5 is inclusive in it and not 65. Above this interval we have 65—69 which has lower limit 64·5 and not 65. This interval contains 3 scores. Our class-interval is of 5 units. Now we know upto 64·5 we have 18 scores and we want only 2 more to make it 20.

For this we enter in the class-interval 65—69 or which we can say 64·5—69·5. 3 scores are lying in this interval and we want 2 out of it. Length of the class-interval = 5

3 covers the length of 5 units

$$\therefore\quad 2\quad ,,\quad ,,\quad ,,\quad ,,\ \frac{5}{3}\times 2 \text{ units}$$

$$=\frac{10}{3}=3{\cdot}33 \text{ units.}$$

Thus upto 64·5 lies 18 scores. To make it 20 we will go upto 64·5 + 3·33 = 67·83 and upto this point will lie 20 scores. Thus 67·83 is called the median.

Mathematically median

$$=\begin{array}{l}\text{Lower limit}\\ \text{of interval} +\\ \text{in which } \frac{1}{2}\\ \text{the total}\\ \text{number of}\\ \text{scores lie.}\end{array}\left\{\frac{\dfrac{\text{total No. of scores}}{2}-\begin{array}{l}\text{sum of all}\\ \text{scores}\\ \text{below that}\\ \text{interval}\end{array}}{\begin{array}{c}\text{frequency which lies in the}\\ \text{interval in which median lies.}\end{array}}\right\}\times\begin{array}{c}\text{length}\\ \text{of}\\ \text{class-}\\ \text{interval}\end{array}$$

Thus in present solution we see 20 scores lie in interval 65—69 and thus median should lie in this interval. Its lower limit is 64·5

Total number of scores = 40

Sum of scores below this interval = 18

Frequency in this interval = 3

Length of the class-interval = 5

$$\therefore\quad Md=64{\cdot}5+\frac{\left(\frac{40}{2}-18\right)}{3}\times 5$$

$$=64{\cdot}5+\frac{2}{3}\times 5=64{\cdot}5+3{\cdot}33$$

$$=67{\cdot}83.$$

Thus this mathematical formula facilitates our calculation work.

Calculation of the median when the frequency distribution contains gaps

Sometimes when the frequency distribution contains gaps, i.e., zero frequency in one or more intervals then it becomes a bit complicated to calculate the median. The method to be followed in such cases is shown in table 5 (a). In this case N=20 and so $\frac{N}{2}=10$. If we count the frequency towards up the column the median is at 24·5, i.e., the lower limit of the interval 25—29 or the upper limit of the interval 20—24. If we count down the frequency column the median falls at 34·5, i.e., the lower limit of the interval 35—39 or the upper limit of the interval 30—34. This discrepancy between these two values of the median is due to two intervals 25—29 and 30—34 each of which has zero frequency. To obtain the median at the same point; whether computed from the bottom or the top of the frequency distribution, include the interval 25—29 in the interval 20—24 which thus becomes 20—29 with a frequency 0+1=1 included in it. Similarly include the interval 30—34 in the interval 35—39 which thus becomes 30—39 with frequency 0+5=5 included in it.

TABLE 5 (a)

Calculation of the median when there are gaps in the distribution.

Class-intervals	*f*	
45—49	3	
40—44	2	
35—39	5	30—40
30—34	0	
25—29	0	20—30
20—24	1	
15—19	6	
10—14	0	
5—9	3	

$$N=20$$

$$\frac{N}{2}=10$$

$$Md=29{\cdot}5+\frac{0}{5}\times 5=29{\cdot}5$$

Thus both the intervals 20—29 and 30—39 have been lengthened from five to ten units to eliminate zero frequency on the adjecent intervals by spreading the numerical frequencies over them. Now if we go up counting ten frequencies the median falls at 29·5 the upper limit of the interval 20—29. Also counting downwards the frequency column ten scores we arrive at a median value 29·5, i.e., the lower limit of the interval 30—39. Thus computation of median from both the ends of the frequency distribution gives the same result with 29·5 as the median.

MODE (Mo)

The mode in a series of scores is that score which occurs most frequently or which occurs maximum number of times.

Mode when scores are ungrouped

Suppose our scores are as 2, 5, 6, 1, 3, 3, 11, 2, 3, 9, 7, 3, 28, 29, 52, 3, 3, 21, 27.

Here we see 3 occurs maximum number of times and thus 3 is the mode. This is the case when scores are ungrouped.

Mode when scores are grouped

When the scores are grouped mode is said to be the middle point of that class-interval which contains the maximum frequency. Thus in table (5) 10 is the maximum frequency which is contained by the class-interval 60—64. The lower limit of it is 59·5 and upper limit is 64·5.

$$\therefore \text{ Middle value}=59{\cdot}5+\frac{(64{\cdot}5-59{\cdot}5)}{2}$$

$$=59{\cdot}5+2{\cdot}5$$

$$=62$$

And this 62 is called the mode,

$$\therefore \qquad Mo = 62$$

It is really very difficult to find out the correct value of the mode. For approximate results we can use the following formula.

$$\text{Mode} = 3\ (\text{median}) - 2\ (\text{mean})$$

$$Mo = 3\ Md - 2M$$

Geometric Mean and Harmonic Mean

Although geometric mean and harmonic mean are very rarely used in statistics but it would be worth while to know something about these central tendencies.

Geometric Mean. If we have the scores 2, 5, 8, 10, 12 then their geometric mean is mathematically.

$$G = \sqrt[5]{2 \times 5 \times 8 \times 10 \times 12} \text{ or } (2 \times 5 \times 8 \times 10 \times 12)^{\frac{1}{5}}$$

Thus it is the fifth root of the product of the scores. In general if there are N scores as $X_1, X_2, X_3 \ldots\ldots X_N$

Their Geometrical Mean will be ${}^{N}\sqrt{X_1, X_1, X_3 \ldots\ldots X_N}$

That is the Nth root of the product of all the scores.

Harmonic Mean. The Harmonic Mean is given by the following mathematical relation.

$$\frac{1}{H} = \frac{1}{N}\Sigma\left(\frac{1}{X}\right)$$

Here H = Harmonic Mean

N = Total number of Scores

X = Score

Suppose we have to find the Harmonic Mean of following numbers.

2, 4, 5, 8, 10, 3.

$$\frac{1}{H} = \frac{1}{6}\left[\frac{1}{2} + \frac{1}{4} + \frac{1}{5} + \frac{1}{8} + \frac{1}{10} + \frac{1}{3}\right]$$

We can easily solve this and get the value of H,

$$H = 3{\cdot}9.$$

Cumulative Frequency. The following table contains two types of frequencies. First one of which is the ordinary frequency and second one f' which is cumulative frequency.

TABLE NO. 6

Class-interval C.I.	Frequency f	Cumulative frequency f'
70–79	2	30
60–69	8	28
50–59	7	20
40–49	5	13
30–39	3	8
20–29	3	5
10–19	2	2

$N=30$

For getting cumulative frequency in the lowest interval same frequency is placed ; in our case it is 2. Then to get the cumulative frequency in the class-interval 20–29 we add the frequencies in it and below it. Thus in present case it will be $3+2=5$.

Similarly cumulative frequency in the interval 30–39 is $3+3+2=8$ and so on.

One thing we should note here that the cumulative frequency in the upper most class-interval is same as total number of ordinary frequencies.

Some times when we are given the data in terms of cumulative frequencies and not in ordinary frequency we should first change cumulative frequency table to ordinary frequency table to get the mean, median etc.

Say as in Table 6 if we are given only cumulative frequencies and we are to find out the mean we change it to ordinary frequency table as below.

Class-interval C.I.	Cumulative frequency f	Ordinary frequency f
70—79	30	30—28=2
60—69	28	28—20=8
50—59	20	20—13=7
40—49	13	13— 8=5
30—39	8	8—5=3
20—29	5	5— 2=3
10—19	2	2=2

Again for this purpose in lowest class-interval we put the same frequency as in above case it is 2. To find ordinary frequency in the class-interval 20—29 we subtract the cumulative frequency 2 from the cumulative frequency above it. In the present case it is 5−2=3.

Similarly to find the ordinary frequency of class-interval 30—39 we subtract cumulative frequency of the class-interval below it from the cumulative frequency of this class-interval.

So it is 8 — 5=3 and so on.

When ordinary frequency is known we can apply any method for finding out the Mean.

When to use the various measures of central tendency

Mean. (1) When the distribution is normal.

(2) When each score should have equal weight in determining the central tendency. Since the mean is the sum of the scores divided by their number each score has equal weight in its determination.

(3) When the S. D. and product—moment coefficients of correlation are to be subsequently computed.

(4) Mean is the most accurate measure of the central tendency.

Median. (1) Median should be used when we want to compute the central tendency quickly and easily.

(2) Sometimes it gives a better notion of the general run of the measures than does the average. For example, the average age of a school class may be unduly raised because the class includes two or three pupils much older than all the rest. They cannot very well be omitted in calculating the average, but if the median is found, they no longer exert a disproportionate effect upon the result.

(3) When the distribution is badly skewed.

(4) When the distribution is not complete.

Mode. (1) When the most often recurring score is sought.

(2) When quickest of the central tendency is needed.

(3) When the most typical case is to be sought.

SUMMARY

(1) Mean for ungrouped scores $(M)=\frac{\Sigma X}{N}$ where Σ is the sum
and X stands for score
and N = total number of scores

(2) Mean for grouped scores by long method $(M)=\frac{\Sigma(fx)}{N}$ where f stands for frequency
x for mid-value of class-interval
and N for total number of scores

(3) Mean for grouped scores by short method $(M)=AM+ci$ where AM = assumed mean
i = length of the class-interval
ci = correction in terms of scores

$ci=\frac{\Sigma fx'}{N}$ where f = frequency

x' = Deviation of the mid-points of different class-intervals from assumed mean.

N = Total number of frequencies

(4) Median (M*d*)

$$=\begin{array}{l}\text{Lower limit of}\\ \text{the interval in}\\ \text{which } \tfrac{1}{2} \text{ the}\\ \text{total No. of}\\ \text{scores lie}\end{array} + \left\{ \frac{\dfrac{\text{Total No. of scores}}{2} - \begin{array}{c}\text{sum of}\\ \text{all scores}\\ \text{below that}\\ \text{interval}\end{array}}{\begin{array}{c}\text{frequency which lies in the}\\ \text{interval in which median lies}\end{array}} \right\} \times \begin{array}{l}\text{Length}\\ \text{of the}\\ \text{class}\\ \text{interval}\end{array}$$

(5) Mode = 3 (Median) − 2 (Mean)

$\text{Mo} = 3\,\text{M}d - 2\text{M}$

(6) Geometric mean $= G = \sqrt[N]{X_1\, X_2\, X_3\, X_4\, X_5 \ldots\ldots X_N}$

(7) The formula for Harmonic mean (H) is :

$\frac{1}{H}=\frac{1}{N}\Sigma\left(\frac{1}{X}\right)$ where H = Harmonic Mean

N = total No. of score

X = Score.

EXERCISES

1. Find the mean of the following numbers :

(a) without further grouping, (b) grouping the numbers by fives, and (c) grouping by tens.

40, 43, 43, 46, 46, 46, 54, 56, 59, 62, 64, 64, 66, 66, 67, 67, 68, 68, 69, 69, 69, 71, 75, 75, 76, 76, 78, 80, 82, 82, 82, 82, 82, 83, 84, 86, 88, 90, 90, 91, 91, 92, 95, 102, 127.

Ans. Mean = 73·2 in all the three cases.

2. Find out the Arithmetic mean for the following table :

Class-interval	Frequency
12·5—17·5	2
17·5—22·5	22
22·5—27·5	19
27·5—32·5	14
32·5—37·5	3
37·5—42·5	4
42·5—47·5	6
47·5—52·5	1
52·5—57·5	1

Ans. Mean = 27·85

3. Find the mean and median for the following data :

Class-interval	Frequency
130—134	7
125—129	6
120—124	5
115—119	2
110—114	3
105—109	1
100—104	6

Hints—AM = 117
$ci = 2{\cdot}5$
Mean = 119·5
Median (M*d*) = 122·5

The median lies in the interval 120—124 whose lower limit is 119·5.

4. Change the ordinary frequency (table 5) into cumulative frequency table and vice versa :

5. Compute the median of the following 10 scores :

Class-intervals C.I.	Frequency *f*
20—21	2
18—19	1
16—17	0
14—15	0
12—13	2
10—11	0
8—9	0
6—7	2
4—5	1
2—3	1
0—1	1
	N = 10

Ans. Median = 9·5

6. Compute the median of the following sixteen scores :

Scores	f
20—21	2
18—19	2
16—17	0
14—15	4
12—13	0
10—11	0
8—9	4
6—7	0
4—5	0
2—3	0
0—1	4
	N=16

Ans. Md=11·5

7. Calculate the median and the mode for the following frequency distribution :

C.I.	f
100—109	1
90—99	0
80—89	1
70—79	9
60—69	23
50—59	32
40—49	36
30—39	7
20—29	6
	N=125

Ans. Md=51·59
Mo=44·5

✓8. Calculate the Mean, Median and Mode of the scores of 185 boys in an achievement test given below :

C.I.	f
70—79	1
60—69	10
50—59	28
40—49	51
30—39	56
20—29	39
10—19	0
	N=185

Ans. Mean (M)=40
Median (Md)=39·1
Mode (Mo)=34·5

CHAPTER 16

MEASURES OF VARIABILITY

In the previous chapter we discussed about the measures of central tendency and pointed out that central tendency is a point on the scale which represents as far as possible the whole distribution. Our next step for describing the distribution completely is to find some measure of the variability of the scores which indicates how far the distribution scatters above and below the central tendency.

We can see the utility of the measures of variability by a simple numerical example. Suppose in a school there are two sections for class X, each including 25 boys. The boys of class X [A] secures marks in English ranging from 20 to 100 and suppose the boys of class X [B] secures marks ranging from 40 to 80. The mean score of the boys of both the sections is 60. But by looking into the means alone we cannot say that the performance of both the classes is same. In the former case the scores are more variable, i.e., separated on a wider range than in the latter case.

If the group is homogeneous, that is made up of individuals of nearly the same ability most of the scores will fall around the same point on the scale, the range will be comparatively short and the variability will be small.

This has been illustrated in the following figure. Here the two frequency distributions are of the same area and the same mean 60, but of different variability.

Group A ranges from 20 to 100 and group B ranges from 40 to 80. That means group A covers two times of the distance than that B covers, though in both the cases central tendency is same.

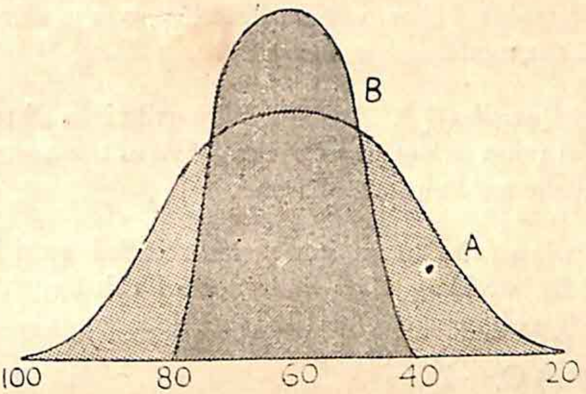


Figure showing two distributions of the same area and same mean 60 but different variability.

The following are the four methods of indicating variability :

(1) Range (R)
(2) The Quartile Deviation (Q)
(3) The Mean Deviation (MD) or Average Deviation (AD)
(4) The Standard Deviation (S.D.).

Range. Range is the difference between the highest and the lowest scores in a series. It is not a satisfactory indicator of variability because it takes into account the extremes of the series only. It is unreliable when N is small or when many gaps [zero f's] occur in the frequency distribution.

Consider the following series :

(a) 100, 40, 40, 20
(b) 80, 30, 50, 40

	central tendency	variability
Series (a)	mean = 50	Range = 100 − 20 = 80
Series (b)	mean = 50	Range = 80 − 30 = 50

Mean being the same in both the series, range in the former case is greater than in the latter. This indicates the scores of the former series are more hetrogeneous.

Quartile Deviation (Q). Before we may discuss Quartile deviation we should be familiar with two terms (*i*) First Quartile (*ii*) Third Quartile.

First Quartile (Q_1). By First Quartile in a distribution we mean that point below which lies 25% of the scores, i.e., one fourth of the total number of scores.

Third Quartile (Q_3). By Third Quartile in a distribution we mean that point below which lies 75% of the scores i.e. three-fourth of the total number of scores.

Quartile deviation is some times called semi-interquartile range. By dividing the difference of third quartile and first quartile by two, we get Quartile Deviation.

In short $Q=\frac{Q_3-Q_1}{2}$

For finding out Quartile Deviation we firstly find out Q_1 and Q_3 and then get the value of Q by using the above formula. It is very easy to get Q_1 and Q_3. These can be calculated in the same way as median. The difference being, in the case of median we choose the class interval in which half of the total number of scores lie. For finding Q_1 we take that class interval below which lies one-fourth of the total number of scores. We take the lower limit of this interval. Suppose we have to find out Q_1 from (table 5). We see total number of scores is 40 and so one-fourth of this is 10. Upto the class-interval 55—59 lies only 8 scores. To make it 10 we are to add 2 more. We enter in the interval 60—64, whose lower limit is 59·5. The length of the class-interval is 5 and in the class-interval 60—64 lies 10 scores. We want 2 out of them.

10 scores cover 5 units

2 ,, ,, $\frac{5\times2}{10}=1$ unit

Thus we add 59·5 and 1 which is 60·5

Thus 10 scores lie at 60·5, which we call the First Quartile (Q_1).

DATA FROM (TABLE 5)

Find Q_1, Q_3 and then Q

	Class-interval	Frequency
	100—104	4
	95—99	2
	90—94	2
Q_3	85—89	5
	80—84	3
	75—79	2
	70—74	1
	65—69	3
Q_1	60—64	10
	55—59	2
	50—54	2
	45—49	2
	40—44	2
		N=40

We can calculate Q_1 by using a formula as in the case of median with a slight change as follows.

$$Q_1 = \begin{array}{l}\text{lower limit}\\ \text{of interval}\\ \text{in which one}\\ \text{fourth of the}\\ \text{total number}\\ \text{of scores lie}\end{array} + \left\{ \frac{\dfrac{\text{Total No. of scores}}{4} - \begin{array}{c}\text{sum of all}\\ \text{the scores}\\ \text{below that}\\ \text{interval}\end{array}}{\begin{array}{c}\text{frequency in the interval in}\\ \text{which } Q_1 \text{ lies}\end{array}} \right\} \times \begin{array}{l}\text{Length}\\ \text{of class}\\ \text{interval}\end{array}$$

In this case

$$Q_1 = 59{\cdot}5 + \frac{\left\{\frac{40}{4} - 8\right\}}{10} \times 5$$

$$Q_1 = 59{\cdot}5 + \frac{2 \times 5}{10} = 59{\cdot}5 + 1 = 60{\cdot}5.$$

After finding out Q_1 we find Q_3. The method is same as above but in this case firstly we take that class-interval below which 3/4th of the total number of scores lie. We take its lower limit. Say in the present case total number of scores are 40 and its 3/4th is 30. We see 27 scores lie upto 80—84. We want 30 scores. So we enter in the class interval 85—89 whose lower limit is 84·5. This interval contains 5 frequencies. Length of the class interval is 5.

5 frequencies or scores cover 5 units

3 frequencies or scores cover 3 units

We add this 3 to 84·5 which is 87·5 and this is the value of Q_3.

In short Q_3=

$$\text{Lower limit of class interval in which 3/4th of total No. of scores lie} + \left\{ \frac{\frac{3 \times \text{total No. of scores}}{4} - \text{sum of all scores below that interval}}{\text{Frequency in the interval in which } Q_3 \text{ lies}} \right\} \times \text{Length of the class-interval}$$

$$Q_3 = 84{\cdot}5 + \frac{[3/4 \times 40 - 27] \times 5}{5}$$

$$= 84{\cdot}5 + 3 = 87{\cdot}5$$

Calculation of Q—We know $Q = \frac{Q_3 - Q_1}{2}$

$$Q_3 = 87{\cdot}5$$

$$Q_1 = 60{\cdot}5$$

$$\therefore \quad Q \text{ (Quartile Deviation)} = \frac{87{\cdot}5 - 60{\cdot}5}{2}$$

$$= \frac{27}{2} = 13{\cdot}5$$

We can say that median is the second Quartile because 50% scores lie below it and therefore rest 50% scores must also lie above it. The readers may compare the formula used for finding the median, First Quartile and Third Quartile. In the case of the median we use $\frac{1}{2}$; In case of First Quartile $\frac{1}{4}$ and in case of Third Quartile $\frac{3}{4}$.

AVERAGE DEVIATION (AD)
OR MEAN DEVIATION (MD)

Average Deviation is the Arithmetic mean of all the deviations of different scores from the mean value of the scores without the regard of sign.

A. D. When Scores are Ungrouped

This definition will be clear from the following example. Suppose we have scores 2, 5, 8, 9, 4, 2

The mean of these scores is $\frac{2+5+8+9+4+2}{6} = \frac{30}{6} = 5$

Deviations of Scores from mean is $2-5=-3$

$5-5=0$

$8-5=3$

$9-5=4$

$4-5=-1$

$2-5=-3$

i.e., $-3, 0, 3, 4, -1, -3$ respectively.

Now we add these deviations without the regard of sign. We take them all to be of + ve signs whether they be + ve or − ve.

Sum of deviations $=14$

Total No. of Scores $=6$

$\therefore$ Mean of the deviations $=\frac{14}{6}=\frac{7}{3}=2{\cdot}33$

M. D. or A. D. $=\frac{\Sigma\,|\,d\,|}{N}$ Where d stands for deviations of scores from mean value of the scores. N = Total number of Scores. | | is the sign of mod., which is an abbreviation of modulus. Whatever value, whether + ve or − ve, we put inside this sign, it gives the magnitude of that value without the regard of sign.

Average deviation (AD) when scores are grouped. "In the case of ungrouped scores the deviations are calculated by subtracting the mean of the scores from individual scores. In grouped scores we have no individual score and in this case we take the middle point of each class-interval and from this subtract the mean of the scores. For this illustration let us take the data from table 2 (b).

Data from Table 2 (b)

METHOD FOR FINDING A. D.

Class-interval	Mid. value	Frequency	Deviation of mid-values from mean		
C.I.	x	f	fx	d	fd
100—104	102	1	102	31	31
95—99	97	2	194	26	52
90—94	92	2	184	21	42
85—89	87	1	87	16	16
80—84	82	2	164	11	22
75—79	77	2	154	6	12
70—74	72	1	72	1	1
65—69	67	1	67	−4	−4
60—64	62	0	0	−9	0
55—59	57	2	114	−14	−28
50—54	52	2	104	−19	−38
45—49	47	2	94	−24	−48
40—44	42	2	84	−29	−58
		$N=20$	$\Sigma fx=1420$		$\Sigma\mid fd\mid=352$

The first four columns remain the same as discussed for finding out the mean by Long method. Of course we can find out the means by Short method as well. To fill the fifth column we find out the deviation of middle values of the class-intervals from the mean.

$$M=\frac{\Sigma(fx)}{N}=71 \text{ (as calculated before)}$$

Mid. value of class-interval 40—44 is 42

$d=$(Mid. value − Mean) $=42-71=-29$

Mid. value of class-interval 80—84 is 82

$d=$(Mid. value—mean) $=82-71=11$

Similarly we can find out the rest of deviations as shown in the above table. Then we multiply each deviation by the frequency in that interval and put it under column fd at the proper place.

Say in the class-interval 40 – 44 frequency = 2

$$\text{and} \qquad d = -29$$

$$\therefore \qquad fd = -58$$

Similarly in the class-interval 80 – 84 frequency = 2

$$\text{and} \qquad d = 11$$

$$\therefore \qquad fd = 22$$

and so on.

Then we add all *fd*s without the regard of signs and call it $\Sigma\,|fd|$ where | | again stands for mod. When we add whole of the *fd* column without the regard of sign we get $\Sigma\,|fd|$ as 352

Now $$AD = \frac{\Sigma\,|fd|}{N} = \frac{352}{20} = 17{\cdot}6$$

where N is total number of frequencies. Thus we see average deviation is calculated by dividing all *fd*s by the total number of frequencies.

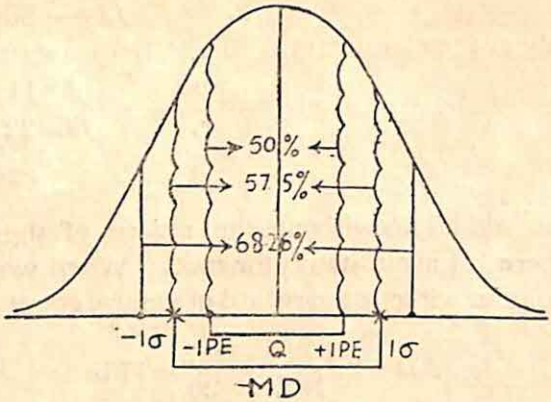
THE STANDARD DEVIATION [SD] OR (σ)

Standard Deviation is the measure of variability which is most important from statistical point of view and is mostly employed in research work. In the case of Mean Deviation or Average Deviation we treat all the deviations without the regard of signs but in calculating the S.D. we square all the deviations separately ; find their sum ; divide the sum by the total number of scores and then find the square root of the mean of the squared deviations. S.D. is also called the root-mean-squared-deviation. The difficulty of sign which is felt in the case of Average Deviation is avoided in the case of Standard Deviation.

The Standard Deviation is less affected by sampling errors than is the Q or the MD and is a more stable measure of dispersion.

In a Normal Distribution the S.D. when measured off above and below the mean, marks the limit of the middle 68·26% (roughly the middle two-thirds) of the distribution, where as M.D. marks the limit of the middle 57·57 per cent of the measures and Q marks the limit of middle 50 per cent.

That is why Q is also called the probable error.[1]



Normal Probability curve showing area covered by S D. ; M.D and Q or P.E.

The standard deviation is the square root of the arithmetic mean of the squares of all the deviations, deviations being measured from the arithmetic mean of the observations.

SD in Ungrouped Scores. Just as in the previous case for finding AD from data of ungrouped scores we calculated the deviations of each score from the mean of the scores the same thing we do for calculating the S.D. for that data. The deviations were $-3, 0, 3, 4, -1, -3$. Then we square each of the deviations and add these squared deviations. Squaring -3 we get 9.

Squaring 0 we get 0
Squaring 3 we get 9

and so on.

Sum of the squared deviations$=9+0+9+16+1+9=44$

Now we divide the sum of the squared deviations by the total number of scores which gives the mean of the squared deviations.

Here mean of squared deviations$=\frac{44}{6}$

To get standard deviation we find out the square root of the 'mean of the squared deviations'.

---

[1] For details see the Chapter '*Normal Probability Curve*' of this book.

Thus standard deviation $=\sqrt{\frac{44}{6}}=2\cdot7$

In general if the sum of the squared deviations $=\Sigma d^2$ and total number of frequencies $=$ N

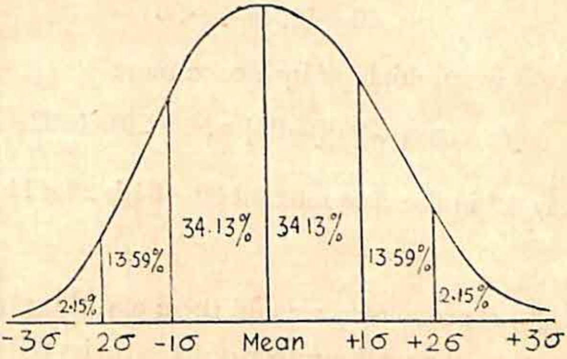
Then standard deviation $=\sqrt{\frac{\Sigma(d)^2}{N}}$

This standard deviation is represented by a Greek letter σ (sigma)

$$\text{Thus } \sigma=\sqrt{\frac{\Sigma(d)^2}{N}}$$

In the case of AD we add the deviations without the regard of sign. In the case of SD the question of sign does not arise because here we add the squares of the deviations and in squaring all signs become positive.

The following figure has been given here just to show the per cent of distribution included in each of the first three σs. The distribution taken here is a normal distribution.



The per cent of the distribution included in each of the first three σs.

Just as in the case of mean there are two methods for calculating it similarly in the case of standard deviation also we use two methods—(1) Long method and (2) Short method.

Calculation of S.D. from Grouped Data (Long Method)

To find out the S.D. in the case of grouped data we firstly find out the mean either by long method, or by short method.

Then the middle value of each class-interval is taken and the deviation of middle value of each class-interval from the mean of the scores is calculated. This we put down under column d as shown below. Then we square each deviation 'd'. As in the case of table 2 (b), d in the class-interval 40 – 44 is – 29 and so we square it to get d^2 which is 841. Similarly other deviations like – 24, – 19, – 14, – 9, – 4, 1, 6, 11, 16, 21, 26, 31 are squared. Then we multiply each squared deviation by the respective frequency to get fd^2. Squared deviation in the class-interval 40 – 44 is 841 and we put it under d^2 column. Frequency in this interval is 2. To get fd^2 we multiply 2 by 841 and put it under fd^2 column. Similarly fd^2 in other cases is found. To facilitate the calculation we firstly calculate d and fd as in the case of A.D. All the columns remain the same up to this step. Then to get fd^2 we multiply fd by d once more.

Generally we adopt this method. For example in the class-interval.

$$40-44, fd=-58$$

To get fd^2 we multiply fd by d once more

$$fd \times d = [-58][-29] = 58 \times 29 = 1682$$

Similarly fd^2 in the class-interval 80 – 84 is $22 \times 11 = 242$

and so on.

When all fd^2 are known we add them and this gives us Σfd^2. Dividing Σfd^2 by the total number of frequencies (N) we get $\frac{\Sigma fd^2}{N}$. The square root of $\frac{\Sigma fd^2}{N}$ is the standard deviation.

In the present case $\Sigma fd^2 = 7780$ Mean = 71

Total number of frequencies = 20

$$\sigma \text{ or S.D} = \sqrt{\frac{\Sigma fd^2}{N}} = \sqrt{\frac{7780}{20}} = \sqrt{389} = 19{\cdot}7$$

Data from Table 2 (b)

(Table showing the calculation of S.D. by the long method)

C.I.	x	f	fx	d	fd	fd^2
100—104	102	1	102	31	31	$31 \times 31 = 961$
95—99	97	2	194	26	52	$26 \times 52 = 1352$
90—94	92	2	184	21	42	$21 \times 42 = 882$
85—89	87	1	87	16	16	$16 \times 16 = 256$
80—84	82	2	164	11	22	$11 \times 22 = 242$
75—79	77	2	154	6	12	$6 \times 12 = 72$
70—74	72	1	72	1	1	$1 \times 1 = 1$
65—69	67	1	67	−4	−4	$(-4) \times (-4) = 16$
60—64	62	0	0	−9	0	$(-9) \times (0) = 0$
55—59	57	2	114	−14	−28	$(-14) \times (-28) = 392$
50—54	52	2	104	−19	−38	$(-19) \times (-38) = 772$
45—49	47	2	94	−24	−48	$(-24) \times (-48) = 1152$
40—44	42	2	84	−29	−58	$(-29) \times (-58) = 1682$
		$N=20$	$\Sigma fx=1420$			$\Sigma fd^2=7780$

$$\text{S.D. or } \sigma = \sqrt{\frac{\Sigma fd^2}{N}} = \sqrt{\frac{7780}{20}} = 19{\cdot}7.$$

Calculation of S.D. from Grouped Data (Short Method)

In the case of long method for finding out S.D. we have seen the complications of calculation. If middle values of the class-intervals are decimal numbers then this method becomes all the more complicated. In short method for finding out S. D. we firstly assume the mean in the same way as we did in the case of finding the mean by short method. Then we find out the middle values of all the class-intervals. The next step is to find out the deviations of mid. values of the class-intervals from assumed mean. These deviations are put down under column X' as shown below. We must note that we can put the deviations either in units of class-intervals or in units of score but former one is preferable. Then we find out the product of each deviation and frequency in that class-interval. This we put down under column fX''. Upto this step our work remains the same as discussed undershort method of finding mean.

Data from Table 2 (b)

(*Table showing the calculation of S.D. by the short method*)

Class-intervals C.I.	Mid. value X	f	Deviation X'	fX'	$(fX')(X')$ $=fX'^2$
100—104	102	1	6	6	$6\times6=36$
95—99	97	2	5	10	$10\times5=50$
90—94	92	2	4	8	$8\times4=32$
85—89	87	1	3	3	$3\times3=9$
80—84	82	2	2	4	$4\times2=8$
75—79	77	2	1	2	$2\times1=2$
70—74	72	1	0	0	$0\times0=0$
65—69	67	1	−1	−1	$(-1)\times(-1)=1$
60—64	62	0	−2	0	$(0)\times(-2)=0$
55—59	57	2	−3	−6	$(-6)\times(-3)=18$
50—54	52	2	−4	−8	$(-8)\times(-4)=32$
45—49	47	2	−5	−10	$(-10)\times(-5)=50$
40—44	42	2	−6	−12	$(-12)\times(-6)=72$

A.M.$=72$ $N=20$ $\Sigma fX'=-4$ $\Sigma fX'^2=310$

$$i=5\ ;\ c=\frac{\Sigma fX'}{N}=\frac{-4}{20}=\frac{-1}{5}$$

$$\text{S.D. or } \sigma=\sqrt{\frac{\Sigma fX^{2}}{N}-c^2}\times i=\sqrt{\frac{310}{20}-\left(\frac{-1}{5}\right)^2}\times5=\sqrt{15\cdot46}\times5$$

$$=3\cdot932\times5=19\cdot7.$$

In this case we assumed the mean as the mid. value of class-interval 70—74, i.e., 72. Then in the X column we put down the mid. values (42, 47, 52, 57 etc.) of all the clase-intervals. In column f we have frequencies of different class-intervals. X' column contains deviations of mid. values of the class-intervals from the assumed mean. Say assumed mean is 72 and mid.-value of class-interval 40—44 is 42.

$\therefore$ X' (Deviation)$=42-72=-30.$

But this is in terms of scores and our class-interval is of 5 scores. Thus we can say $X'=\frac{-30}{5}=-6$ intervals. This point we have discussed before as well.

Similarly other deviations in unit of class-intervals are calculated.

Then to find out fX' we multiply each X' by frequency in the class-interval. Frequency in the class-interval 40—44 is 2 and $X'=-6$.

Thus $fX'=2\times(-6)=-12$ and so on.

Now our formula for finding out S.D. by this method is

S. D. or $\sigma=\sqrt{\frac{\Sigma f X'^2}{N}-c^2}\times i$ where f=frequency

X'=deviation of mid.-value of class-interval from assumed mean

N=Total No. of frequencies

c=Correction in terms of class-interval

i=Length of the class-interval

We know correction in terms of class-interval is equal to $\frac{\Sigma(fX')}{N}$. This has been explained earlier.

Here $$c=\frac{-4}{20}=\frac{-1}{5}$$

because $$\Sigma(fX')=-4 \text{ and } N=20.$$

Now we want $\Sigma(fX'^2)$. We multiply each fX' by the corresponding X' of each class-interval and put it respectively under column fX'^2.

Say fX' in the class-interval 40—44 is (−12). We multiply it by X' i.e. −6 and get 72. This is fX'^2. Similarly fX'^2 in other cases is found. The sum of all fX'^2 gives $\Sigma(fX'^2)$. In the present case it comes out to be 310.

i is the length of the class-interval. Here it is 5

$$\therefore \quad \sigma=\sqrt{\frac{\Sigma(fX'^2)}{N}-c^2}\times i=\sqrt{\frac{310}{20}-\left(\frac{-1}{5}\right)^2}\times 5=\sqrt{15{\cdot}46}\times 5$$

$$=(3{\cdot}9325\times)=19{\cdot}66=19{\cdot}7.$$

Thus Standard Deviation=σ=19·7 (The same result we got by long method).

Interpretation of S. D.

In the above question S.D.=19·7 and the mean (M) is 71. So $\frac{2}{3}$ of individuals should be expected between (71+19·7) and (71−19·7), i. e., between (Mean+1σ) and (Mean−1σ) as discussed in the chapter 'Normal probability curve'.

Thus between 90·7 and 51·3 there should lie $\frac{2}{3}$ of the total number of cases.

$\frac{2}{3}$ of 20=13 (approximately).

So approximately 13 individuals should lie between 90·7 and 51·3. These two figures include nine complete intervals from 50—54 to 90—94, and from the table 2 (b) we see these intervals actually contains 13 individuals.

Checking the solution of S. D.

(1) As shown above if the percentage of cases that lie between $+1\sigma$ and -1σ is approximately $\frac{2}{3}$ of the total number of case (N) i.e., 68% then our calculation of σ is correct. This check is applicable only in the case of a normal distribution.

(2) Another rough check is to compare the S.D. obtained with the total range of measurements. In large samples ($N=500$ or more) the S.D. is about $\frac{1}{6}$ of the total range or stated otherwise, the total range is about 6σ.

In smaller samples the ratio of range to S.D becomes smaller as indicated in the table below.

[Table showing the ratios of the total range to the S.D. in a distribution for different values of N.]

N	$\frac{\text{Range}}{\sigma}$	N	$\frac{\text{Range}}{\sigma}$	N	$\frac{\text{Range}}{\sigma}$
5	2·3	40	4·3	400	5·9
10	3·1	50	4·5	500	6·1
15	3·5	100	5·0	700	6·3
20	3·7	200	5·5	1000	6·5

(3) A third method of checking the calculation for S.D. has been illustrated below with the help of the example given under table 2 (b).

In this method we take the assumed mean (A.M.) at one interval above or below that class-interval in which the assumed mean is taken for calculating σ. As in the previous case for calculating σ by short method we computed $\Sigma X'$, $\Sigma(fX')$ and $\Sigma(fX'^2)$ similarly by taking another assumed mean we compute $\Sigma x_1'$, $\Sigma fX_1'$ and $\Sigma(fX_1'^2)$ where X_1' denotes the deviation of mid.-value of the class-interval from the second assumed mean in terms of class-interval.

Let
$$A=\Sigma(fX'^2)$$
$$A'=\Sigma fX_1'$$
$$B=\Sigma(fX'^2)$$

$$B' = \Sigma(fX_1'^2)$$

If $$A - A' = N$$

and $A + A' = B - B'$ then our calculation is correct.

[*Example showing the check of computation work for calculating S. D.*]

Data from Table 2 (b)

Class-interval C.I.	f	(x')	(fx')	(fx'^2)	(x_1')	(fx_1')	$(fx_1'^2)$
100—104	1	6	6	36	7	7	49
95—99	2	5	10	50	6	12	72
90—94	2	4	8	32	5	10	50
85—89	1	3	3	9	4	4	16
80—84	2	2	4	8	3	6	18
75—79	2	1	2	2	2	4	8
70—74	1	0	0	0	1	1	1
65—69	1	−1	−1	1	0	0	0
60—64	0	−2	0	0	−1	0	0
55—59	2	−3	−6	18	−2	−4	8
50—54	2	−4	−8	32	−3	−6	18
45—49	2	−5	−10	50	−4	−8	32
40—44	2	−6	−12	72	−5	−10	50
	N=20		$\Sigma(fx') = -4$	$\Sigma(fx'^2) = 310$		$\Sigma(fx_1') = 16$	$\Sigma(fx_1'^2) = 322$

In this example $A=-4$ and $A'=16$

$\therefore$ $A-A'=-4-16=-20=N$ ...(1)

Again $B=310$ and $B'=322$

$\therefore$ $B-B'=310-322=-12$

$A+A'=-4+16=12$

$\therefore$ $B-B'=A+A'$ (signs are not considered)...(2)

In both the cases (1) and (2) we are getting the difference in minus sign on both the sides of the equation which is immaterial because we are to see the magnitude and not the sign. To avoid this confusion we can write our check formula as $A'-A=N$ and $B'-B=A'+A$

Applying these in the above case

$16-(-4)=20=N$

and $322-310=-4+16=12$

Some Examples for Calculating the S.D.—Below are given two more examples for calculating the S.D. by the short method.

Example 1

Calculation of the mean and the standard deviation of the scores made by 160 pupils of the second year science class of Intermediate colleges (U. P.) on the achievement test in general science.[2]

C.I.	f	x'	fx'	fx'^2
60−69	6	3	18	54
50−59	29	2	58	116
40−49	45	1	45	45
30−39	55	0	0	0
20−29	22	−1	−22	22
10−19	3	−2	−6	12
	N=160		$\Sigma fx'=93$	$\Sigma fx'^2=249$

A.M. (Assumed Mean)$=34{\cdot}5$

$$\text{M (Mean)}=34{\cdot}5+\frac{93}{160}\times 10=40{\cdot}3$$

$$\text{S. D. (Standard Deviation)}=\sqrt{\frac{249}{160}-\frac{93}{160}\times\frac{93}{160}}\times 10$$

$$=11{\cdot}04$$

---

[2] This data has been taken from the author's thesis.

Example 2

Data from table 3

Calculation of the standard deviation (S.D.) of the scores obtained by 125 pupils upon an achievement test in general science.

C.I.	f	x'	fx'	fx'^2
100—109	1	5	5	25
90—99	0	4	0	0
80—89	1	3	3	9
70—79	9	2	18	36
60—69	23	1	23	23
50—59	32	0	0	0
40—49	36	-1	-36	36
30—39	17	-2	-34	68
20—29	6	-3	-18	54

$N=125 \qquad \Sigma fx'=-39 \quad \Sigma fx'^2=251$

$$\text{S. D. (Standard deviation)}=\sqrt{\frac{\Sigma fx'^2}{N}+\left(\frac{\Sigma fx'}{N}\right)^2}\times i$$

$$=\frac{10}{125}\sqrt{251\times125-39\times39}$$

$$=171{\cdot}3\times\frac{10}{125}$$

$$=13{\cdot}7.$$

Calculation of σ from Original Scores

The Short-Method for computing σ can also be applied in case of ungrouped scores as shown below. The assumed mean is taken at zero and each score becomes a deviation (X′) from the A.M., that is each score (X) remains unchanged. The correction (C) is the difference between the actual mean (M) and assumed mean (A.M.)

Thus $$C=\frac{\Sigma X'}{N}=M-0=M$$

The mean can be calculated as before by summing the scores and dividing it by N. To find σ we square X′ and then calculate $\Sigma X'^2$, divide it by N, subtract C^2 or M^2 from

$\frac{\Sigma X'^2}{N}$. The square root of $\left(\frac{\Sigma X'^2}{N}-M^2\right)$ gives σ

Thus S.D. or $\sigma=\sqrt{\frac{\Sigma X'^2}{N}-M^2}$ or $\sqrt{\frac{\Sigma(X'^2}{N}-\left(\frac{\Sigma X'}{N}\right)^2}$

X′ has the same value as X

So we can write $\sigma=\sqrt{\frac{\Sigma(X^2)}{N}-\left(\frac{\Sigma X}{N}\right)^2}$

$$=\frac{\sqrt{N\Sigma(X^2)-(\Sigma X)^2}}{N}$$

This method of calculating σ is useful when there are relatively few scores, say fifty or less, and when the scores are expressed in not more than two digits so that squaring of the numbers may not become a complicated task. The calculating machine facilitates the work.

[To illustrate the Calculation of the σ from original scores when the assumed mean is taken at zero and the data is ungrouped.]

Scores X	X′ or X	X'^2 or X^2
8	8	64
10	10	100
15	15	225
17	17	289
8	8	64
25	25	625
13	13	169
12	12	144
16	16	256
6	6	36
N=10	$\Sigma X'=130=\Sigma X$	$\Sigma X^2=1972$

$$\text{A.M.}=0$$

$$M=\frac{\Sigma X'}{N} \text{ or } \frac{\Sigma X}{N}=\frac{130}{10}=13$$

$$C=\frac{\Sigma X'}{N} \text{ or } C=M-0$$

$$C=\frac{130}{10}=13 \text{ or } C=13-0=13$$

$$C^2=169$$

$$\sigma=\frac{\sqrt{N\Sigma(X^2)-(\Sigma X)^2}}{N}=\frac{\sqrt{(10)(1972)-(130)^2}}{10}$$

$$=\frac{\sqrt{19720-16900}}{10}=\frac{\sqrt{2820}}{10}=5{\cdot}31$$

We may find the same result if we calculate σ by the method discussed previously for ungrouped scores.

Here Scores are 8, 10, 15, 17, 8, 25, 13, 12, 16, 6

Their sum=130.

$N=10$

∴ Mean=M=13

Deviations d are 5, 3, 2, 4, 5, 12, 0, 1, 3, 7

d^2 are 25, 9, 4, 16, 25, 144, 0, 1, 9, 49

$\Sigma d^2=282$

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}}=\sqrt{\frac{282}{10}}=\frac{\sqrt{2820}}{10}=5{\cdot}31$$

Here we see that the former method is easier in this case.

Effect upon σ of (a) Adding a constant to each score (b) Multiplying each score by the same number

(a) If a constant is added to each given score the mean is increased by the amount of the constant added but the variability (σ) is not affected. If a constant is subtracted from each score, the mean is decreased by the amount of the constant, and σ remains unchanged.

This can be illustrated by the following example where each score has been increased by 5.

Original Scores (X)	d_1	d_1^2	$(X+5)$	d_2	d_2^2
6	2	4	$6+5=11$	2	4
5	1	1	$5+5=10$	1	1
4	0	0	$4+5=9$	0	0
3	-1	1	$3+5=8$	-1	1
2	-2	4	$2+5=7$	-2	4
$5\lfloor\underline{20}$ 4	$\Sigma d_1^2=10$		$5\lfloor\underline{45}$ 9	$\Sigma d_2^2=10$	

$M_1=4$ $\qquad\qquad$ $M_2=\frac{45}{5}=9$

$\sigma_1=\sqrt{\frac{\Sigma d_1^2}{N}}=\sqrt{\frac{10}{5}}=1{\cdot}41$ $\qquad\qquad$ $\sigma_2=\sqrt{\frac{10}{5}}=1{\cdot}41$

Here we see M is increased by 5 but σ remained constant on adding 5 to each score.

(b) If each given score is multiplied by a constant, the net effect of this operation is to multiply the mean and σ by that constant number as shown below where each score is multiplied by 10.

Original Scores (X)	d_1	d_1^2	(X) (10)	d_2	d_2^2
6	2	4	60	20	400
5	1	1	50	10	100
4	0	0	40	0	0
3	-1	1	30	-10	100
2	-2	4	20	-20	400
$5\lfloor\underline{20}$ 4		$\Sigma d_1^2=10$	$5\lfloor\underline{200}$ 40		$\Sigma d_2^2=1000$
$M_1=4$			$M_2=40$		
$\sigma_1=\sqrt{\frac{10}{5}}=1{\cdot}41$			$\sigma_2=\sqrt{\frac{1000}{5}}=\sqrt{200}$ $=14{\cdot}1$		

Here we see M is increased 10 times and σ has also increased 10 times on multiplying each score by 10.

When to use different measures of variability

(a) Use the Range. (1) When the data are too scant or too scattered to justify the calculation of any other measure of variability.

(2) For a quick, inspectional measure of variability.

(b) Use the Q. (1) When there are scattered or extreme cases, that is when the distributions are truncated.

(2) When the median is the measure of central tendency adopted. Q should be used as a measure of variability.

(c) Use the M.D. or A.D. (1) It is occasionally employed as an index of the extent to which the measures are spread out on either side of the mean.

(2) A.D. or M.D. is used when it is desired to weight all the deviations according to their size.

(d) Use the S.D. (1) When coefficients of correlation or measures of reliability are subsequently to be computed.

(2) When highest degree of reliability is wanted.

(3) When more weight to extreme deviations from the mean is to be given because it is derived from the squared deviations.

THE COEFFICIENT OF VARIATION

Sometimes the tester needs to compare the variability of a group upon different tests or he desires to compare the variabilities of different groups upon the same test. The Quartile Deviation, Average Deviation and Standard Deviation are not appropriate measures for comparing such variabilities. These measures tell the dispersion of test scores around their means in terms of the units of the test.

But due to the differences in units of the test scores we cannot compare the variability of a group upon different tests. Again we cannot compare the variability of different groups upon the same test because of the difference in their means. Coefficient of variation or V is a measure which is independent of the units of test scores and takes account of both the central tendency and the variability of the group.

So for measuring the variability we use V.

The formula for calculating V is : $V=\frac{100\times\sigma}{M}$. It is also called the coefficient of relative variability. The following example will illustrate the calculation of the value of V.

Example. In a school the mean height of the ten year old boys is 50 inches and $\sigma=3$; their mean weight is 60 pounds and $\sigma=6$ pounds. In which trait is that group of boys more variable ?

Solution. Coefficient of variation for height

$$=V_{ht}=\frac{100\times 3}{50}=6$$

Coefficient of variation for weight

$$=V_{wt}=\frac{100\times 6}{60}=10$$

This evidently shows that the boys are 6/10 or 60% as variable in height as in weight.

Example. In a type writing competition a group of boys completes an average of 35 words in the first minute, σ being 5. At the end of the second minute same group completes an average of 50 words, σ being 8. Compare the relative dispersion of the two groups in different periods.

Solution. On comparing the σs of the two distributions we may conclude that the group was nearly $1\frac{1}{2}$ times as variable at the end of the second minute as it was at the end of the first minute. But to compare the relative dispersion we should consider the fact that the means have also increased from 35 to 50 when σ increases from 5 to 8.

For one minute period $V=\frac{100\times 5}{35}=14{\cdot}5$

For two minute period $V=\frac{100\times 8}{50}=16$

Thus instead of being about 66% as variable in one minute period as in two, the group is about 90% as variable when the mean score is considered as well as the variability.

SUMMARY

(1) Q=Quartile Deviation$=\frac{Q_3-Q_1}{2}$

where Q_1=first quartile and Q_3=third quartile.

(2) Average Deviation AD in case of ungrouped scores$=\frac{\Sigma\,|\,d\,|}{N}$ where d=deviation of scores from the mean

$|\;|$=mod.

N=Total number of scores.

(3) Average Deviation AD in the case of grouped scores $=\frac{\Sigma\,|\,fd\,|}{N}$ where d=deviations of mid-values of class-interval from the mean

$|\;|$=mod.

N=number of scores

f=frequency

(4) SD$=\sigma$=Standard Deviation

when scores are ungrouped $\sigma=\sqrt{\frac{\Sigma d^2}{N}}$

where d and N has the same meaning as above

(5) σ in case of grouped scores$=\sqrt{\frac{\Sigma fd^2}{N}}$

This is the formula for long method.

Here f, d, N has same meaning as above.

(6) σ in case of grouped scores (Short Method)

$$=\sqrt{\frac{\Sigma fx^2}{N}-c^2}\times i$$

where x'=Deviations of mid. values of the class interval from assumed mean in unit of class-interval.

f=frequency

N=Total number of scores

c=correction$=\frac{\Sigma fx'}{N}$

i=length of the class-interval.

EXERCISES

1. Calculate the mean and SD from the following data :

Mid.-values	Frequency
2·5	2
7·5	5
12·5	7
17·5	13
22·5	21
27·5	16
32·5	8
37·5	3

Ans. Mean = 21·9 (App.)
SD = 8·0 (App.)

2. Find out the SD of the following numbers
(a) without further grouping (b) Grouping the numbers by fives (40—, 45—, 50,— etc.) (c) Grouping by tens (40—, 50—, etc.)

40, 43, 43, 46, 46, 46, 54, 56, 59, 62, 64, 66, 66, 67, 67, 68, 68, 69, 69, 69, 71, 75, 75, 76, 76, 78, 80, 82, 82, 82, 82, 82, 83, 84, 86, 88, 90, 90, 91, 91, 92, 95, 102, 127.

Ans. (a) $\sigma = 17{\cdot}3$
(b) $\sigma = 17{\cdot}5$
(c) $\sigma = 18$

3. Calculate the Mean and SD from Data [Table 3 (b)] of this book.

Ans. Mean = 24·75
S.D. = 9·6 (App.)

4. Calculate quartile deviation Q and the standard deviation σ, from the following data grouped into frequency distribution.

Class-interval	Frequency
195—199	1
190—194	2
185—189	4
180—184	5
175—179	8
170—174	10
165—169	6
160—164	4
155—159	4
150—154	2
145—149	3
140—144	1

Ans. $Q = 8{\cdot}28$

$\sigma = 12{\cdot}63$

Note—The students should use both the Long and Short Methods of finding σ in above exercises and compare the results.

5. Calculate the mean and the standard deviation of the scores obtained by 108 pupils of the first year science class of intermediate colleges (U.P.) on the achievement test in general science.

C.I.	*f*
50—59	2
40—49	19
30—39	33
20—29	40
10—19	14
	N=108

Ans. Mean=30

Standard deviation=9·8

6. Calculate the mean, median and standard deviation for the following frequency distribution :

C.I.	*f*
100—109	1
90—99	1
80—89	4
70—79	14
60—69	35
50—59	26
40—49	30
30—39	12
20—29	2
	N=125

Ans. Mean=56·26

Median=56·61

Standard deviation=14·538

CHAPTER 17

CORRELATION

OUR PREVIOUS CHAPTERS DEALT WITH TWO KINDS OF measures, first the averages or measures of central tendencies and second the measures of variability. In this chapter we will discuss the third type of measure.

Meaning of Correlation

If we have two series of variables and if we know the relation between them, then we can predict the future effect of one variable on the other. Say if the number of teachers in a school vary according to the number of students we can any time predict the number of teachers with the knowledge of the number of students. Whenever two variables are so related that the increase or decrease in one corresponds to the increase or decrease in another or conversely increase or decrease in one corresponds to the decrease or increase in another, they are called to be correlated. The method by which the relation between the two correlated variables is calculated is known as *method of correlation*.

To have a clear idea of correlation we should take the following example in which scores of 6 boys A, B, C, D, E, F, in Science and Maths Examinations have been given.

(a) Boys	Scores (Science)	Scores (Maths.)
A	25	35
B	22	30
C	20	26
D	18	25
E	17	23
F	16	18

Here we see one boy who is good in Science is also so in Maths. A is getting the highest marks in Science and also in Maths. B is next to highest in Science and also in Maths. Thus we see the marks of each boy in Science and Maths are related.

Perfect Positive Correlation

As in the above example A is best in Science and also best in Mathematics.

B is next best in Science and next best in Mathematics and so on.

Thus A ranks first in both the subjects.
B ranks second in both the subjects
C ranks third in both the subjects
and so on.

Here we see one variable is perfectly correlated with another variable. Another thing which we observe is that the increase of marks of an individual in one subject corresponds to the increase of marks in another subject or the decrease of marks of an individual in one corresponds to the decrease of marks in another subject. Such correlation is called positive correlation. This is the example of perfect positive correlation.

Perfect Negative Correlation

Let the scores of six boys in Science and Maths be as below.

(b) Boys	Science	Mathematics
A	25	18
B	22	23
C	20	25
D	18	26
E	17	30
F	16	35

In this example we see A is best in Science but worst in Maths.
B is next best in Science but above worst in Maths and so on.
Thus A ranks first in Science but sixth in Maths.
B ranks second in Science but fifth in Maths and so on.

Of course here the relations between the scores are perfect. In the case of the perfect correlation if we know the rank of an individual in one trait we can at once judge its rank in another trait by seeing the scores.

In this example we observe that the increase of marks of an individual in one subject corresponds to the decrease of marks in another. Thus here variables have inverse relation.

Such correlation is termed as negative correlation. Since this is again the example of perfect correlation we call it perfect negative correlation.

Zero Agreement or No Correlation

If the two variables are distributed so that half of the highest values of one variable correspond with the similar half of the highest values of the other, and the reverse is true for the remaining values, the correlation is neither positive nor negative and the values are uncorrelated ; that is, the relationship is such as may be due to chance.

An agreement approaches neutrality, when the variables are not correlated positively or negatively, it becomes less and less significant, since the closer the agreement comes to neutrality, the greater is the indication that is due to chance.

Coefficient of Correlation

The statistical method where by relationship is expressed on a quantitative scale is called the 'coefficient of correlation' and is designated by the letter r.

Thus in case (a) above, correlation is perfect positive because there is one to one correspondence between the variables. In such a case we say coefficient of correlation is $+1$.

In case (b) we see negative or inverse relationship is perfect and its coefficient of correlation is -1.

In case of zero agreement there is no relation between the variables and so the coefficient of correlation is zero.

Thus we observe that correlation can range from perfect positive to zero to perfect negative. In other words we can say that coefficient of correlation ranges from $+1 \leftarrow 0 \rightarrow -1$. It is not necessary that the correlation may always be perfect positive or perfect negative. Positive correlation may range from $0 \rightarrow +1$. Similarly negative correlation may range from $0 \rightarrow -1$.

·9, ·8, ·7, ·6, ·5, etc. are the examples of the coefficients of positive correlation and in which case we talk of high positive correlation. Similarly when the coefficients of negative correlation are $-·9$, $-·8$, $-·7$, $-·6$, $-·5$ we say there is high negative correlation. When co-efficients of positive correlation are ·1, ·2, ·3, ·4, etc. we call them as low positive correlation. Similarly $-·1$, $-·2$, $-·3$, $-·4$ are coefficients of low negative correlations.

More accurately we can interpret coefficient of correlation (*r*) as follows :

r from ·00 to ± ·20 means almost negligible correlation.

r from ± ·20 to ± ·40 means low positive or negative correlation depending upon +ve or −ve sign.

r from ± ·40 to ± ·70 means substantial positive or negative correlation depending upon +ve or −ve sign.

r from ± ·70 to ± 1·00 means high to very high positive or negative correlation depending on positive or negative sign.

The Scatter Diagram

To calculate the coefficient of correlation we take two sets of data of the variables. When the number of scores of each set is great then it becomes very difficult to find out *r* without arranging the scores in a tabular form. So for this reason we arrange them in a tabular form as shown below and this is known as the scatter diagram.

How to prepare the Scatter Diagram

Suppose the following two data give the marks obtained by the school boys in English and Hindi respectively. We are to arrange them in a scatter diagram.

Marks in English :—15, 16, 17, 17, 17, 16, 16, 16, 21, 24, 23, 23, 19, 21, 18, 18, 18, 25, 25, 25, 25, 25, 26, 26, 26, 27, 32, 32, 33, 33, 33, 34, 26, 29, 29, 29, 27, 26, 24, 31, 31, 31, 29, 29, 26, 26, 34, 34, 32, 32, 32, 32, 28, 28, 35, 35, 35, 35, 40, 40, 40, 40, 40, 42, 42. 42, 42, 38, 38, 45, 45, 45, 45, 45, 46, 46. 46, 46, 46, 47, 47, 47, 49, 49, 48, 48, 50, 51, 52, 53; 53, 50, 51, 51, 55, 56, 57, 60, 60, 63.

Marks in Hindi :—20, 21, 21, 25, 25, 30, 31, 32, 36, 36, 36, 36, 32, 32, 32, 33, 33, 30, 30, 30. 30, 34, 34, 34, 33, 40, 40, 40, 40, 40, 40, 41, 41, 41, 41, 41, 42, 42, 42, 43, 44, 44, 44, 45, 45, 45, 40, 40, 40, 40, 46, 46, 46, 42, 42, 42, 47, 47, 48, 48, 48, 48, 44, 44, 44, 49, 49, 49, 49, 50, 50, 50, 51, 51, 55, 55, 55, 54, 54, 54, 59, 59, 58, 58, 56, 57, 57, 57, 53, 53, 53, 53 53, 60, 61, 62, 62, 63, 67, 67.

(TABLE 7) SCATTER DIAGRAM

Marks in Hindi (X Variables)

Marks in English (Y Variables)	20—29	30—39	40—49	50—59	60—69	$f(y)$	Row Mean
55—64					(6) //////	6	64·5
45—54				(24) //////// //////// ////////	(1) /	25	54·9
35—44			(15) /////////// ////			15	44·5
25—34		(8) ////////	(29) ////////// ////////// /////////			37	42·3
15—24	(5) /////	(12) ////////////				17	32·1
$f(x)$	5	20	44	24	7	100	
Column mean	19·5	23·5	32·9	49·5	58·07		

To prepare the scatter diagram first we group the scores of both the sets into the class-intervals. A class-interval of ten scores always facilitates the work but it is not a binding rule. For grouping the scores into the class-interval we proceed in the same way as discussed in Chapter II. The class-intervals are then arranged as above. One set of the class-intervals is placed on the left hand of the diagram, the lowest one being placed at the bottom. In this case the class-intervals for the marks in English are 15—24, 25—34 etc. We put them on the left hand side. At the bottom 15—24 is placed and the rest are put in the ascending orders on the same side. The class-intervals of the second set (Hindi Scores) are placed on the top and begin with the lowest one at the left hand corner. Thus in our case they are 20—29, 30—39, etc. We call marks in English as '*y*' variables and marks in Hindi as '*x*' variables. Then we see the marks of individual boy in English and Hindi. Say the first boy is getting 15 marks in English and 20 in Hindi. We put a 'tally' in the first cell from the left of the bottom row. Second boy obtains 16 marks in English and 21 in Hindi. Again 'tally' is placed in the same cell. This cell represents the marks of those who obtain 15—24 marks in English and 20—29 marks in Hindi.

Thus a boy getting 27 marks in English and 40 marks in Hindi will be represented by a tally in a cell No. 3 from the left in the last but the bottom row. Similarly we fill up the whole diagram by putting the tallies. As shown in the above diagram, if we add the tallies in the rows we get the frequency of the class-interval against that row.

Thus frequency in the class-interval 55—64 is 6.

The frequency in the class-interval 45—54 is 25 and so on.

Similarly if we add the tallies in the column we get the frequency in the class-interval against each column.

Thus frequency in the class-interval 20—29 is 5.

The frequency in the class-interval 30—39 is 20 and so on.

On adding the frequencies of the different rows and the different columns separately we get the total as 100 in each case. If we do not get the same total in both the cases it means we have committed some mistake in putting the tallies.

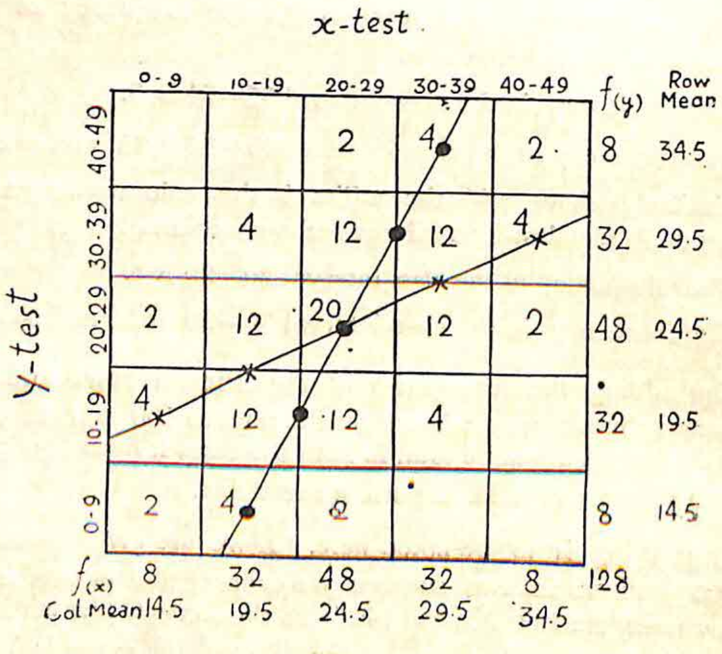
One thing is to be noted here. If we are given the scores of 100 students in two subjects we cannot judge at random as to how many students are good in both the subjects but with the help of the scatter diagram we can do so. For example in the above diagram we see 6 boys are getting marks between 60—69 in

Hindi and between 55—64 in English and these are the best boys of the class. Similarly 5 boys are getting marks between 15—24 in English and between 20—29 in Hindi. These are the worst type of boys. 15 boys, who are at the centre of the diagram are of medium category. Again this scatter diagram is helpful for us in calculating the coefficient of correlation by product-moment method.

The Graphic representation of the correlation coefficient

It may be seen through the graph as to how correlation coefficient measures relationship. The following figures [Table 7 (a) ; 7 (b) ; 7 (c) and 7 (d)] depict correlation of ·50 ; 1·00 ; ·00 and —·75 respectively. It has been explained below as to how these tables denote graphically the relationship between two test scores, namely x—test and y—test.

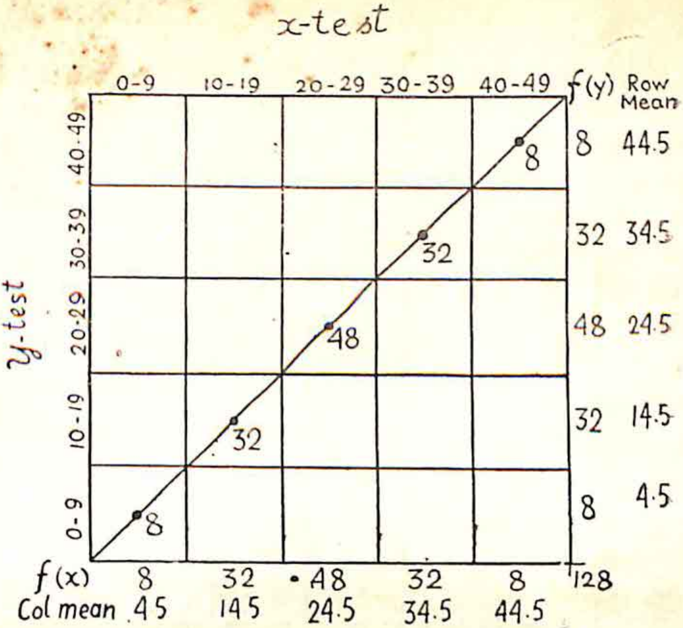
TABLE 7 (a)



y-test \ x-test	0-9	10-19	20-29	30-39	40-49	$f_{(y)}$	Row Mean
40-49			2	4	2	8	34.5
30-39		4	12	12	4	32	29.5
20-29	2	12	20	12	2	48	24.5
10-19	4	12	12	4		32	19.5
0-9	2	4	2			8	14.5
$f_{(x)}$	8	32	48	32	8	128	
Col Mean	14.5	19.5	24.5	29.5	34.5		

$r = ·50$

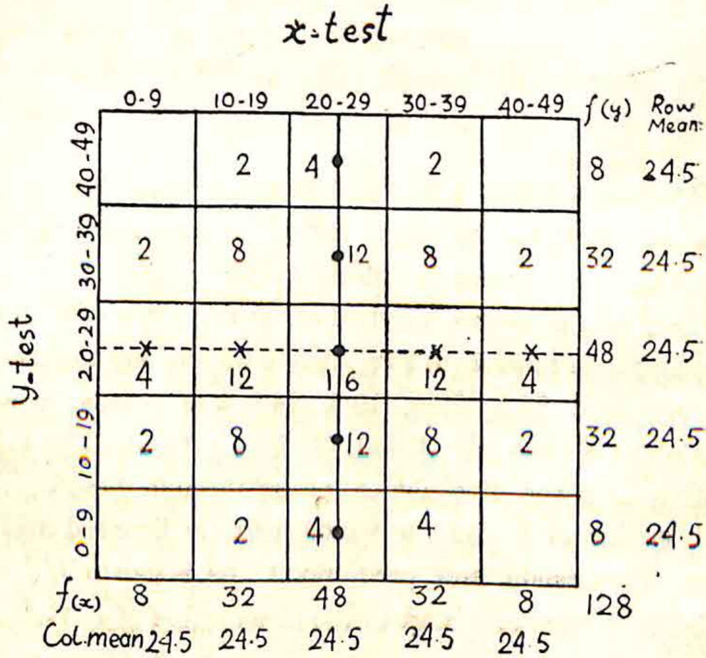
TABLE 7 (b)



y-test	0-9	10-19	20-29	30-39	40-49	f(y)	Row Mean
40-49					8	8	44.5
30-39				32		32	34.5
20-29			48			48	24.5
10-19		32				32	14.5
0-9	8					8	4.5
f(x)	8	32	48	32	8	128	
Col mean	4.5	14.5	24.5	34.5	44.5		

$r = 1.00$

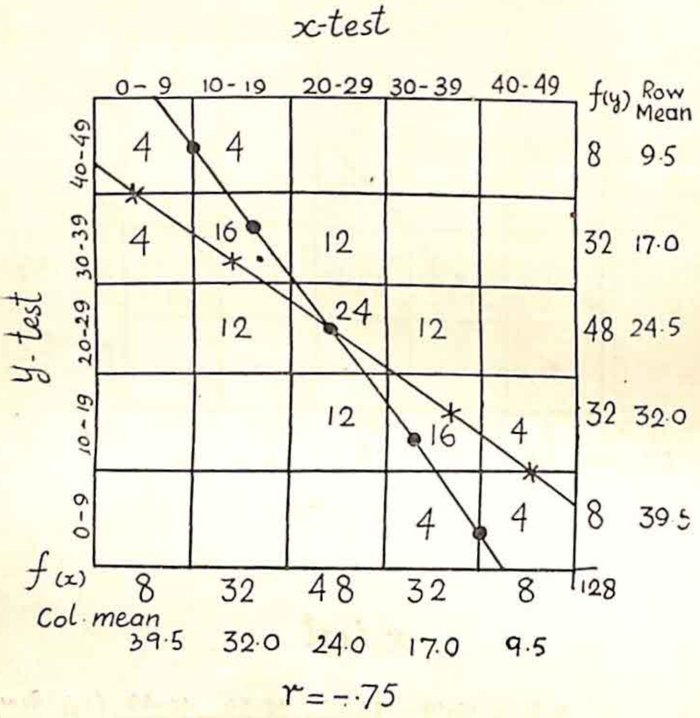
TABLE 7 (c)



y-test	0-9	10-19	20-29	30-39	40-49	f(y)	Row Mean
40-49		2	4	2		8	24.5
30-39	2	8	12	8	2	32	24.5
20-29	4	12	16	12	4	48	24.5
10-19	2	8	12	8	2	32	24.5
0-9		2	4	4		8	24.5
f(x)	8	32	48	32	8	128	
Col. mean	24.5	24.5	24.5	24.5	24.5		

$r = .00$

TABLE 7 (d)



x-test

y-test	0 - 9	10 - 19	20-29	30-39	40-49	f(y)	Row Mean
40-49	4	4				8	9.5
30-39	4	16	12			32	17.0
20-29		12	24	12		48	24.5
10-19			12	16	4	32	32.0
0-9				4	4	8	39.5
f(x)	8	32	48	32	8	128	
Col. mean	39.5	32.0	24.0	17.0	9.5		

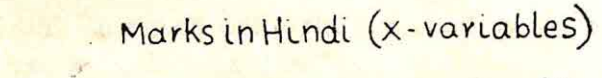
$r = -.75$

The scores along the top of these tables denote the *x*-test scores and those laid off at the left of the table as *y*—test scores. The means of *x*-columns are placed at the bottom of each table and the means of *y*-rows are laid off at the right of each table. For example in table 7 (a) 14·5, 19·5, 24·5, etc. are the means of *x*-columns and 34·5, 29·5, 19·5, 14·5 are means of *y*-rows. The crosses denote the mean of each *y*-row falling in each *x*-column and the line which passes through these crosses is called a line of regression which will be discussed under the chapter 'Regression and predection.' As is shown in the table this line indicates the change in the mean value of *y* over a given range of *x*. The dots denote the mean of each *x*-column falling in each y-row and the line of regression through these dots denote the change in the mean value of x over a given range of y. These two lines together represent the relationship between the variables x and y. The closer are these two lines of regression the higher is the correlation between the variables. Table 7 (a) shows a correlation of ·50. Here the correlation is positive because the regression lines, through the means of columns and rows, run from the upper right to the lower left hand section of the diagram. The angle between these two regression lines is of 45°. Table 7 (d) represents the correlation of $-·75$. Here the regression lines are more closer than in table 7 (a). The correlation is negative because the regression lines through the means of the columns and the rows run from the upper left to the lower right hand section of the diagram. Table 7 (b) is the graphical representation of the case of perfect positive correlation, i.e., when $r=1·00$. Here the means of columns and rows overlap each other and the two lines of regression form a single line which is shown as a diagonal from the upper right to the lower left of the diagram. Table 7 (c) is the case of zero correlation. Here the two lines of regression are perpendicular to each other. Here it is to be noted that neither the value of mean of y-scores change over the whole range of x and nor the value of mean of x-scores over the whole range of y.

In all the above diagrams the lines of regression pass through all the mean values of y-row and x-column. This is only possible when test-scores are quite accurate and true. In rest cases when the lines of regression are drawn they pass through the majority of the points indicating the mean of rows and of columns. As in the case of the following table the correlation between the marks in Hindi and English in a group of 100 college students have been represented graphically. The row means are 65·5, 54·9, 44·5, 42·3 and 32·1. The column means are 19·5, 23·5, 32·9, 49·5 and 58·07. On plotting these means and joining the successive crosses and dots we get lines jagged in appearance.

Two straight lines have been drawn in to describe the general trend of these irregular lines. These two lines pass as closely as possible through the dots and the crosses. These are the regression lines or the lines of the 'best fit'. In this case the coefficient of correlation is ·918. Hence we see the lines are very close to each other. Again the correlation is positive because the regression lines, through the mean of the columns and rows, run from the upper right to the lower left-hand section of the diagram.

Graphical representation of correlation between the marks in Hindi and English in a group of 100 college students. (Table 7)



	20 29	30-39	40-49	50-59	60-69	f(y)	Row Mean
55-64					6	6	64.5
45-54				24	1	25	54.9
35-44			15			15	44.5
25-34		8	29			37	42.3
15-24	5	12				17	32.1
f(x)	5	20	44	24	7	100	
Col Mean	19.5	23.5	32.9	49.5	58.07		

$r = \cdot 918$

Pearson's Product Moment Method of Calculating the Coeff. of Correlation (r)[1]—

Suppose we are to find out the coefficient of correlation of the marks of 100 students in (Table **7**). We firstly prepare the scatter diagram as above. Marks in Hindi are called as x-variables and marks in English as y—variables. When we add the frequencies in different columns separately it denotes the frequency of the class interval in that column. Thus frequency of the class-interval 20—29 is 5, frequency of the class-interval 30—39 is 20 and so on. These we call as $f_{(x)}$ and put them against $f_{(x)}$. Similarly $f_{(y)}$ represents the frequencies of the different class-intervals in different rows. Thus 6 is the frequency of the class-interval 55—64 etc.

---

[1] The coefficient of correlation, r is called the "Pearson r" after Professor Karl Pearson who developed the product-moment method, following the earlier work of Galton and Bravais. See Walker, H. M., *Studies in the History of Statistical Method* (1929) Chapter 5, pp. 96-111.

(Data from Table 7)

Product moment method of calculating the coefficient of correlation (r)

from the scores of 100 boys in English and Hindi

Mark in Hindi (x-variables)

Marks in English—(y—variables)	20—29	30—39	40—49	50—59	60—69	$f(y)$	y'	$f(_y)y'$	$f(_y)y'^2$	$\Sigma x'y'$
55—64					6 (4) 24	6	2	12	24	24
45—54				24 (1) 24	1 (2) 2	25	1	25	25	26
35—44			15 (0) 0			15	0	0	0	0
25—34		8 (1) 8	29 (0) 0			37	—1	−37	57	8
15—24	5 (4) 20	12 (2) 24				17	—2	−34	68	44
$f(_x)$	5	20	44	24	7	100	−34	154	$\Sigma x'y'=102$	
x'	—2	−1	0	1	2	$[\Sigma f(x)'x'=8]$				
$f(_x)x$	−10	−20	0	24	14	(x)				
$f(_x)x'^2$	20	20	0	24	28	$[\Sigma f(_x)x'^2=92]$				

$$r=\frac{\frac{\Sigma x'y'}{N}-C_xC_y}{\sigma_x\sigma_y}=\frac{\frac{102}{100}-\left(\frac{8}{100}\right)\left(\frac{-34}{100}\right)}{\cdot 956\times 1\cdot 193}=\frac{\frac{102}{100}+\frac{8\times 34}{100\times 100}}{\cdot 956\times 1\cdot 193}$$

$r=\cdot 918.$

Thus in this case there is a very high class positive correlation.

(Probable Error) $\text{PE}_r=\frac{\cdot 6745\ (1-r^2)}{\sqrt{N}}=\frac{\cdot 6745\ (1-\cdot 918^2)}{\sqrt{100}}=\cdot 0106$

The next step to be followed is to assume the mean in both the distributions. In the present case for x distribution we assume the mean in the interval 40—49 and for y distribution in the interval 35–44. We deepen the lines of the diagram in these intervals. The deviation of the mid.-values of the class-interval in both the distributions are taken from their respective assumed mean and put down against x' and under y' respectively. We take the deviations in terms of the class-interval.

Thus 2, 1, 0, —1,–2 are the deviations under y' and
–2, –1, 0, 1, 2 are the deviations under x'
$f_{(y)}y'$ and $f_{(x)}x'$ columns are filled in the same way as before.

These are the products of the frequency and the deviation of each class-interval in both the distributions. Thus the frequency in the interval 55—64 is 6 and deviation y' is 2

$$\therefore\ f_{(y)}y=6\times2=12$$

Similarly $f_{(x)}x'$ is determined in each case. Then we find out $f_{(y)}y'^2$ and $f_{(x)}x'^2$ as is done in the case of SD. For this we simply multiply $f_{(y)}y'$ by y' and $f_{(x)}x'$ by x' respectively.

Total number of scores $=100=$N

Correction for x distribution $=C_x=\frac{8}{100}$ in terms of the class-interval.

Correction for y distribution $=C_y=\frac{-34}{100}$ in terms of the class-interval.

$$\sigma_x=\sqrt{\frac{92}{100}-\frac{8^2}{100^2}}=\cdot956$$ in terms of the class-interval.

Here σ_x is SD of x distribution in terms of the class-interval.

σ_y=SD of y distribution in terms of the class-interval.

$$=\sqrt{\frac{154}{100}-\frac{34^2}{100^2}} \quad =1\cdot193 \text{ in terms of the class-}$$

interval.

Now we step to find $\Sigma x'y'$ which is the sum of the product deviations. Say in the last cell of the upper most row, frequency is 6. The deviation of this cell from the assumed mean of English marks distribution, i.e., y' is 2 intervals and its deviation from the assumed mean of Hindi Score distribution, i.e., x' is 2 intervals. The product of these two deviations x' and y' $=2\times2=4$. This is the product deviation of one frequency in this cell. Therefore, the product deviation of 6 frequencies in this cell $=6\times2\times2=24$. We put (4) above 6 and 24 below 6. Similarly the cell containing the frequency 1 has $x'=2$ and $y'=1$. Therefore, product deviation $=2\times1=2$. This is the product deviation for one frequency. Of course there is only one frequency we put (2) at the top of 1 and 2 at its bottom.

The cell containing the frequency 5 has $x'=-2$ and $y'=-2$

$\therefore$ Product – deviation for one frequency$=(-2)\times(-2)=4$

$\therefore$ Product—deviation for 5 frequencies$=4\times5=20$

Hence we put 20 at the bottom of 5

The cell containing 15 frequencies has $x'=0$ and $y'=0$

$\therefore$ Product-deviation for this cell$=0$.

All the cells bounded by the deep inked lines will have product-deviations zero because in half of the cells $x'=0$ and in the rest half $y'=0$, hence the product is automatically zero. We have accidently no frequency in the second quardant of the diagram. Had there been any frequency in any cell of this quadrant, its product deviation would have been negative because in this case y' will be positive and x' negative. Hence the product will be negative. Similarly product deviation in the fourth quadrant will be negative. But in the first and third quadrant product deviation will be positive. Now to fill $\Sigma x'y'$ column we add all the figures, put down below the frequencies in each cell. The sum of +ve and −ve product deviations is taken which gives $\Sigma x'y'$. Thus in the present case $x'y'$ for 6 frequencies $=24$, $x'y'$ for 25 frequencies in the second row from above $=26$ and so on.

Sum of all the product deviations $=\Sigma x'y'=102$

Now the formula for the coefficient of correlation by the product moment method is $r=\dfrac{\dfrac{\Sigma x'y'}{N}-Cx\ Cy}{\sigma_x\ \sigma_y}$.

Here $\Sigma x'y'$ = Sum of all the product deviations

N = Total No. of frequencies

Cx = Correction for x distributions (taken in unit of class-intervals)

Cy = Correction for y distribution (,,)

σ_x = SD for x distribution (,,)

σ_y = SD for y distribution (,,)

Cx, Cy, σ_x, σ_y are easily found as before but in this case we must note that these should be calculated in terms of the class-interval because we are taking product deviation in terms the class-interval.

Thus in the present case $r=\cdot 918$ which shows that there is a very high class positive correlation. A boy who is good in English is equally good in Hindi. A boy who is poor in English is poor in Hindi too.

Probable Error. In the case of product moment correlation the Probable Error is given by the formula

$$PE_r=\frac{\cdot 6745\ (1-r^2)}{\sqrt{N}}$$

Here PE_r stands for the Probable Error

N = Total number of frequencies

r = Product moment correlation

As shown in the above example $PE_r=\cdot 0106$.

Example—Let us take one more example to illustrate the calculation of Product Moment Coefficient of Correlation. Let us determine the Product Moment Coefficient of Correlation between I. Qs. and General Science Scores of 107 pupils of the fourth form as shown in the following table.

→ I. Q. ←

↑ GENERAL SCIENCE SCORES ↓

	110 – 114	115 – 119	120 – 124	125 – 129	130 – 134	135 – 139	f_y	y'	$f_y y'$	$f_y y'^2$	$x'y'$
110 – 119							0	6	0	0	
100 – 109				$(5)^{1^{5}}$			1	5	5	25	5
90 – 99							0	4	0	0	
80 – 89			$0^{1^{0}}$		$(6)^{1^{6}}$		2	3	6	18	6
70 – 79			$0^{3^{0}}$	$(2)^{1^{2}}$		$(6)^{1^{6}}$	5	2	10	20	8
60 – 69	$(-2)^{1^{-2}}$	$(-3)^{3^{-1}}$	$0^{11^{0}}$	$(3)^{3^{1}}$	$(4)^{2^{2}}$		20	1	20	20	2
50 – 59	$0^{4^{0}}$	$0^{10^{0}}$	$0^{10^{0}}$	$0^{3^{0}}$	$0^{1^{0}}$	$0^{2^{0}}$	30	0	0	0	0
40 – 49	$(6)^{3^{2}}$	$(8)^{8^{1}}$	$0^{18^{0}}$	$(-2)^{2^{-1}}$			31	−1	−31	31	12
30 – 39	$(4)^{1^{4}}$	$(18)^{9^{2}}$	$0^{3^{0}}$	$(-4)^{2^{-2}}$			15	−2	−30	60	18
20 – 29		$(6)^{2^{3}}$	$0^{1^{0}}$				3	−3	−9	27	6
f_x	9	32	47	12	4	3	107		$\Sigma f_y y'$	$\Sigma f_y y'^2$	$\Sigma x'y'$
x'	−2	−1	0	1	2	3			$= -29$	$= 201$	$= 57$
$f_x x'$	−18	−32	0	12	8	9	$\Sigma f(x)x'$	$= -21$			
$f_x x'^2$	36	32	0	12	16	27	$\Sigma f(x)x'^2$	$= 123$			

$$c_x=\frac{-21}{107} \qquad c_y=\frac{-29}{107}$$

$$c_x c_y=\frac{21\times 29}{107\times 107}=\cdot 05$$

$$\text{S. D.}_{(x)}=\sqrt{\frac{123}{107}-\frac{(21)^2}{(107)^2}}=\frac{1}{107}\times 112\cdot 8=1\cdot 05$$

$$\text{S. D.}_{(y)}=\sqrt{\frac{201}{107}-\frac{(29)^2}{(107)^2}}$$

$$=\frac{1}{107}\times 143\cdot 7=1\cdot 3$$

$$r=\frac{\frac{\Sigma x'y'}{\text{N}}-c_x c_y}{\text{S. D.}_{(x)}\times \text{S. D.}_{(y)}}=\frac{\frac{57}{107}-\cdot 05}{1\cdot 05\times 1\cdot 3}=\frac{\cdot 53-\cdot 05}{1\cdot 365}=\cdot 36$$

This is a low positive correlation.

Coefficient of correlation from Ungrouped Data

In the above cases for determining the value of 'r' we took deviations x' and y' from the assumed means of x and y distributions respectively. For this to get the correct value of r we subtracted C_x C_y from $\frac{\Sigma x'y'}{\text{N}}$. If we take the deviation from the actual means then in that case we need not subtract C_x C_y, and our formula for the coefficient of correlation becomes :

$$r=\frac{\Sigma d_1\, d_2}{\text{N}\ \sigma_x \sigma_y}$$

Where d_1 and d_2 are deviations of the scores from the actual means in x and y distribution respectively. But in the case of ungrouped scores we know

$$\sigma_x=\sqrt{\frac{\Sigma d_1^2}{\text{N}}} \text{ and } \sigma_y=\sqrt{\frac{\Sigma d_2^2}{\text{N}}}$$

$$\therefore\quad r=\frac{\frac{\Sigma d_1 d_1}{\text{N}}}{\sqrt{\frac{\Sigma d^2{}_1}{\text{N}}}\sqrt{\frac{\Sigma d^2{}_2}{\text{N}}}}=\frac{\Sigma d_1 d_2}{\sqrt{\Sigma d^2{}_1}\ \sqrt{\Sigma\, d_2^2}}$$

For ungrouped scores when N is not large we will use this formula. An illustration of this method is given in Table 8. The weights and heights of 6 boys are given. We take weights as x variable and heights as y variable. The mean $M_{(x)}$ is calculated by adding all the scores and dividing it by 6.

Similarly $M_{(y)}$ i.e. the mean of the scores of $y-$variables is calculated. $M_{(x)} = \frac{90}{6} = 15$ and $M_{(y)} = \frac{96}{6} = 16$.

d_1 is the deviation of the score under x from the mean 15

d_2 is the deviation of the score under y from the mean 16

Under d_1^2 and d_2^2 we put down the squares of the values of d_1 and d_2 respectively as before.

In $d_1\, d_2$ column the product of d_1 and d_2 is put. The sum of all $d_1\, d_2^{s}$ gives $\Sigma d_1\, d_2$,

Similarly the sum of all d_1^2 gives Σd_1^2 and the sum of all d_2^2 gives Σd_2^2. We put the values in the above formula and get the value of the coefficient of correlation.

TABLE 8

Calculation of (*r*) *from ungrouped scores*

(The weights and heights of 6 Boys being given.)

Boy	Weight	Height	Deviation				
	x	y	d_1	d_2	d_1^2	d_2^2	d_1d_2
A	15	13	0	-3	0	9	0
B	12	12	-3	-4	9	16	12
C	20	16	5	0	25	0	0
D	19	15	4	-1	16	1	-4
E	14	19	-1	3	1	9	-3
F	10	21	-5	5	25	25	-25

$\Sigma d_1^2 = 76\ \ \Sigma d_2^2 = 60 \Sigma d_1 d_2 = 20$

$$M_{(x)}=\frac{90}{6}=15$$

$$M_{(y)}=\frac{96}{6}=16$$

$$r=\frac{\Sigma d_1 d_2}{\sqrt{\Sigma d_1^2\ \Sigma d_2^2}}=\frac{-20}{\sqrt{76\times 60}}=\frac{-20}{675}=-\cdot 3$$

Here we have low negative correlation.

Calculation of (*r*) from the un-grouped scores when the deviation are taken from the assumed means of the scores

As illustrated above the calculation of *r* from ungrouped scores when deviations are taken from the actual means becomes a difficult one with the introduction of decimal numbers. It is by chance that in the above case we did not get the actual mean in decimal number, otherwise all our deviations would have been decimal numbers and it would have been difficult to get their squares. To avoid this difficulty we assume the mean and deviations are taken from it. In both the x and y distributions means AMx and AMy are assumed. Then deviations of individual scores x' and y' are respectively taken from the assumed means AMx and AMy.

Data from Table 8

r From Ungrouped Scores (Deviation are taken from the Assumed Means)

(1) Boys	(2) Weights	(3) Heights	(4)	(5)	(6)	(7)	(8)
	x	y	x'	y'	x'^2	y'^2	$x'y'$
A	15	13	−5	−6	25	36	30
B	12	12	−8	−7	64	49	56
C	20	16	0	−3	0	9	0
D	19	15	−1	−4	1	16	4
E	14	19	−6	0	36	0	0
F	10	21	−10	2	100	4	−20
N=6					$\Sigma x'^2$ $=226$	$\Sigma y'^2$ $=114$	$\Sigma x'y'$ $=70$

$$AM_x = 20 \qquad AM_y = 19.$$

$$M_x = 15 \qquad M_y = 16$$

$$\therefore\ C_x = 15 - 20 = -5 \qquad C_y = 16 - 19 = -3$$

$$(C_x)^2 = 25 \qquad (C_y)^2 = 9$$

$$\sigma x = \sqrt{\frac{226}{6} - 25} \qquad \sigma y = \sqrt{\frac{114}{6} - 9}$$

$$= 3{\cdot}6 \qquad = 3{\cdot}1$$

$$\therefore r = \frac{\frac{\Sigma x'y'}{N} - C_x\, C_y}{\sigma x \times \sigma y}$$

$$= \frac{\frac{70}{6} - 15}{3{\cdot}6 \times 3{\cdot}1} = \frac{-20}{6 \times 3{\cdot}6 \times 3{\cdot}1} = -{\cdot}3$$

In the present case we assumed the mean for x distribution as 20.

So deviations x' from the different scores of this distribution are -5, -8, 0, -1, -6, -10 etc.

Similarly y' are deviations of different scores in y distribution from the assumed mean $AM_{(y)}$ which is 19.

x'^2 column is filled by putting the squares of x' and similarly y'^2 is the square of y'. In column (8) $x'\,y'$ is the product of deviation x' and y'.

Say in the top most row x' is -5 and y' is -6

$\therefore\ x'\,y' = 30.$

Adding all the $x'\,y'$ in this column we get $\Sigma x'y'$.

Now the formula for finding the coefficient of correlation is

$$r = \frac{\frac{\Sigma x'y'}{N} - C_x\, C_y}{\sigma x\ \sigma y}$$

We know M_x = Actual mean

M_x = Assumed mean + correlation

$M_x = AM_x + C_x$

$\therefore\ C_x = M_x - AM_x$

We find out M_x by adding all the scores of this distribution and dividing it by the total number.

AM_x and M_x are thus known to us. We can calculate C_x

Here $M_x = \frac{90}{6} = 15$ and $AM_x = 20$

$\therefore \; C_x = 15 - 20 = -5$

Similarly C_y (correction for y distribution) $= 16 - 19 = -3$

$\Sigma x'y' = 70.$

σ_x is SD for x distribution

$$= \sqrt{\frac{\Sigma x'^2}{N} - (C_x)^2}$$

$$\sigma_y = \sqrt{\frac{\Sigma y'^2}{N} - (C_y)^2}$$

Putting down these values we get $r = -\cdot 3$ (app.)

This is the same result which we got by another method discussed earlier.

RANK DIFFERENCE METHOD OF CORRELATION

(ρ) (Rho)

What is Ranking

Suppose we are testing the students in English and Hindi. If various examiners are to test the same boy, say (A), in English and Maths. they will allot him different marks. One may give 20 marks to A in English and 30 marks in Hindi.

Another may give 16 marks to A in English and 25 marks in Hindi.

A third may give 15 marks to A in English and 28 marks in Hindi and so on.

So the correlation between the marks will not depend upon the marks in English and Hindi only but on the examiner too and thus there can be error in our judgment.

But if we judge the students in English and Hindi in terms of the ability we will say, A gets first rank in Hindi and Second in English. Of course, there can also be divergence in the opinion of the examiners but not so much as in the above case. For this reason we rank the different individuals according to their ability in different traits, such as traits of honesty, beauty and intelligence etc.

If we test five students getting marks as 10, 6, 8, 5, 3, we say first student getting 10 marks stands first rank, third student getting 8 marks stands second rank and the student getting 6 marks stands third rank and so on.

Rank Difference Method of Correlation

If we are given two sets of scores for finding out the coefficient of correlation we firstly rank both the distributions according to the ability or merit or any behaviour. Say as in the following case the marks of eight boys in History and Geography are given. We firstly rank the students according to the marks in History. E getting the highest marks in History stands first rank in that subject, H stands second rank in History, D stands third, B stands fourth and A stands fifth rank. Now F and G gets the same score, i.e., 8. The question arises whether F should be ranked sixth or G should be ranked sixth. It is definite that these two boys are to fill the 6th and the 7th rank. So we rank both of them $\frac{6+7}{2}$ or 6·5. The next man C will be ranked eight. Note that C will not rank seventh because that has already been occupied as discussed before. Similarly we rank the boys in Geography. Here E ranks first. Each of C and D ranks 2·5 and so on.

RANK DIFFERENCE METHOD OF CALCULATING

The Coefficient of Correlation (ρ)

[TABLE 8′]

Boys	Scores in History	Scores in Geog.	Rank in History	Rank in Geog.	Diff. of Rank	Square of Rank Diff.
					D	D_2
A	12	13	5	5	0	0
B	14	18	4	4	0	0
C	6	24	8	2·5	5·5	30·25
D	15	24	3	2·5	·5	·25
E	22	26	1	1	0	0
F	8	12	6·5	6	·5	·25
G	8	6	6·5	8	1·5	2·25
H	17	8	2	7	5	25
	N=8					$\Sigma D^2=58$

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6 \times 58}{8(64-1)}$$

$$= 1 - \frac{6 \times 58}{8 \times 63} = 1 - \cdot 69 = \cdot 31 \text{ (App.)}$$

This is a low positive correlation.

After the boys are arranged in the order of their ranks we find out the difference of their ranks without any consideration of sign and put it under column D. Thus the rank difference

of H is $2-7=-5$ or 5

of $G=6{\cdot}5-8=-1{\cdot}5$ or $1{\cdot}5$ etc.

Then we find out the rank difference square in each case and put it under column D^2. Thus D^2 for H is 5, so D^2 is 25. D for A is zero and thus D^2 is also zero. We add all D^2*s* getting ΣD^2. In the present case the value of ΣD^2 is 58.

Now substituting the values of ΣD^2 and N in the following formula ρ can be calculated.

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

Here $N=8$ and $\Sigma D^2=58$.

Substituting the value of N and ΣD^2 in the above formula the value of the rank difference correlation can be calculated. The result in this case is not accurate and if we want to change ρ in terms of r we can do it with the help of table 8 (a).

The formula for the probable error in the case of rank difference method is

$$PF_r = \frac{\cdot 7063\,(1-r)}{\sqrt{N}}.$$

If we know the value of ρ we can convert it into r by consulting the table 8 (a). On substituting the values of r in the above formula we get the value of PEr. In the present case $\rho = \cdot 31$ and from the table corresponding $r = \cdot 3232$. Again the value of $N=8$.

$$\therefore \; PE_r = \frac{\cdot 7063\,(1-\cdot 3232^2)}{\sqrt{8}}$$

$$= \cdot 22$$

TABLE No. 8 (a)

A table to infer the value of r from any given value of ρ.

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

ρ	r	ρ	r	ρ	r	ρ	r
·01	·0105	·26	·2714	·51	·5277	·76	·7750
·02	·0209	·27	·2818	·52	·5378	·77	·7847
·03	·0314	·28	·2922	·53	·5479	·78	·7943
·04	·0419	·29	·3025	·54	·5580	·79	·8039
·05	·0524	·30	·3129	·55	·5680	·80	·8135
·06	·0628	·31	·3232	·56	·5781	·81	·8230
·07	·0733	·32	·3335	·57	·5881	·82	·8325
·08	·0838	·33	·3439	·58	·5981	·83	·8421
·09	·0942	·34	·3542	·59	·6081	·84	·8516
·10	·1047	·35	·3645	·60	·6180	·85	·8610
·11	·1151	·36	·3748	·61	·6280	·86	·8705
·12	·1256	·37	·3850	·62	·6379	·87	·8799
·13	·1360	·38	·3935	·63	·6478	·88	·8893
·14	·1465	·39	·4056	·64	·6577	·89	·8936
·15	·1569	·40	·4158	·65	·6676	·90	·9080
·16	·1674	·41	·4261	·66	·6775	·91	·9173
·17	·1778	·42	·4363	·67	·6873	·92	·9269
·18	·1882	·43	·4465	·68	·6971	·93	·9359
·19	·1986	·44	·4567	·69	·7069	·94	·9451
·20	·2091	·45	·4609	·70	·7167	·95	·9543
·21	·2195	·46	·4771	·71	·7265	·96	·9635
·22	·2299	·47	·4872	·72	·7363	·97	·9727
·23	·2403	·48	·4973	·73	·7460	·98	·9818
·24	·2507	·49	·5075	·74	·7557	·99	·9909
·25	·2611	·50	·5176	·75	·7654	1·00	1·0000

Disadvantages of the rank difference method

(a) If the value of N is less than 30 then PE r is greater and in that case there are more chances of error. That is why we should use this method of rank difference only, when the number of scores are more than 30.

(b) Here we do not mention particular measures of scores but the ranks. Say scores of three boys are 80, 50, 10 ; we rank them as 1st, 2nd and 3rd. But if we see the actual numerical difference between the first and the second, it is 30. Difference between the 3rd and the 4th is 40.

In doing so we are entertaining error by using difference method.

Thus for accurate results we should not use this method.

Coefficient of correlation by the method of gains (R)

This method is almost the same as the rank difference method. Here again we mark the score or behaviour of an individual according to its rank. Thus if we take the same example as in the case of rank difference, the first four columns, i.e., scores in History, scores in Geog. rank in Hist. and rank in Geog. remains the same.

Gain method of calculating the coefficient of correlation

Data from Table 8

(1) Boys	(2) Scores in History	(3) Scores in Geog.	(4) Rank in History	(5) Rank in Geog.	(6) Gain of (5) over (4) G	(7) Gain of (4) over (5) G′
A	12	13	5	5	0	0
B	14	18	4	4	0	0
C	6	24	8	2·5	5·5	...
D	15	24	3	2·5	·5	...
E	22	26	1	1	0	0
F	8	12	6·5	6	·5	...
G	8	6	6·5	8	—	1·5
H	17	8	2	7	...	5

$\Sigma G = 6{\cdot}5$ $\qquad$ $\Sigma G' = 6{\cdot}5$

In this case columns (1), (2), (3), (4), (5) are same as in the rank difference method.

$$R=1-\frac{6\ \Sigma G}{N^2-1}; \qquad \Sigma G=6\cdot5$$
$$N=8$$

$$\therefore\ R=1-\frac{6\times6\cdot5}{63}$$

$$=\cdot38.$$

Here columns (6) and (7) are named as G and G′. Say C is ranked 2·5th in Geog. in column (5) and 8th in history in column (4). So he is gaining in his rank in Geog. over to history by 8−2·5=5·5. We put it under column G. Similarly D gains in Geog. over history and G=3—2·5=·5 etc.

If we take the rank of H in Geog. and history we see that the gain of H in history over Geog. is 7−2=5 and we put it under column G′. Thus G′ and G′ columns are filled. In the case of A and B none gains in their ranks hence G and G′ are both zero. We add all G and all G′ separately giving ΣG and $\Sigma G'$ respectively. Note both ΣG and $\Sigma G'$ are equal. If they are not equal it means calculation is wrong.

The formula for finding the coefficient of correlation by this method is.

$$R=1-\frac{6\Sigma G}{N^2-1}$$ Where ΣG is the sum of the gains

N=total Number of frequencies.

Substituting these values of ΣG and N in the above formula we get the value of R. With the help of the following table we convert R in terms of *r* (Product moment correlation).

This method is also not accurate for the same reason as discussed under Rank-difference method of correlation. The value obtained by this method is even less accurate than by rank difference method. Hence we try to avoid this method.

TABLE No. 8 (b)

A table to infer the value of r from any given value of R

R	*r*	R	*r*	R	*r*	R	*r*
·00	·000	·26	·429	·51	·742	·76	·937
·01	·018	·27	·444	·52	·753	·77	·942
·02	·036	·28	·458	·53	·763	·78	·947
·03	·054	·29	·472	·54	·772	·79	·952
·04	·071	·30	·486	·55	·782	·80	·956
·05	·089	·31	·500	·56	·791	·81	·961
·06	·107	·32	·514	·57	·801	·82	·965
·07	·124	·33	·528	·58	·810	·83	·968
·08	·141	·34	·541	·59	·818	·84	·972
·09	·158	·35	·554	·60	·827	·85	·975
·10	·176	·36	·567	·61	·836	·86	·979
·11	·192	·37	·580	·62	·844	·87	·981
·12	·209	·38	·593	·63	·852	·88	·984
·13	·226	·39	·606	·64	·860	·89	·987
·14	·242	·40	·618	·65	·867	·90	·999
·15	·259	·41	·630	·66	·875	·91	·991
·16	·275	·42	·642	·67	·882	·92	·993
·17	·291	·43	·654	·68	·889	·93	·995
·18	·307	·44	·666	·69	·896	·94	·996
·19	·323	·45	·677	·70	·902	·95	·997
·20	·338	·46	·689	·71	·908	·96	·998
·21	·354	·47	·700	·72	·915	·97	·999
·22	·369	·48	·711	·73	·921	·98	·9996
·23	·384	·49	·721	·74	·926	·99	·9999
·24	·399	·50	·732	·75	·932	1·00	1·0000
·25	·414						

SUMMARY

(1) Co-efficient of correlation by the product-moment method when scores are grouped is $r = \dfrac{\dfrac{\Sigma x'y'}{N} - C_x C_y}{\sigma_x \sigma_y}$

Here $\Sigma x'y'$ is the sum of the product of the deviations from their assumed means in x and y distributions respectively.

Cx = correction (x distribution)
Cy = correction (y distribution)
σx = SD for (x distribution)
σy = SD for (y distribution)
All in units of the class-intervals.

(2) The Co-efficient of correlation in the case of unground scores, deviation d_1 and d_2 being taken from the actual means of x and y distributions respectively is

$$\frac{\Sigma d_1 d_2}{\sqrt{\Sigma d_1^2}\sqrt{\Sigma d_2^2}}$$

(3) Co-efficient of correlation in the case of undergrouped scores, deviations x' and y' being taken from the assumed means AMx and AMy respectively in x and y distributions respectively.

$$r = \frac{\dfrac{\Sigma x'y'}{N} - C_x C_y}{\sigma_x \sigma_y}$$

(4) Rank difference correlation $(\rho) = 1 - \dfrac{6\Sigma D^2}{N(N^2-1)}$

where D = difference of the ranks

(5) Co-efficient of correlation by gain method

$$(R) = 1 - \frac{6\Sigma G}{(N^2-1)}$$

where G stands for gain.

(6) Probable error in the case of product moment method

$(PEr) = \dfrac{\cdot 6745\,(1-r^2)}{\sqrt{N}}$ where r = Product moment correlation.

(7) Probable error in the case of rank difference method

$(PEr) = \dfrac{\cdot 7063\,(1-r^2)}{\sqrt{N}}$ where r = Product moment correlation.

EXERCISES

1. Calculate the value of r from the following table :

Wt. in pounds (x-variables)

Height in inches (y-variables)	100—109	110—119	120—129	130—139	140—149	150—159	160—169	170—179
72—73								1
70—71			1	3	3	4	2	3
68—69			4	11	6	3	2	2
66—67		2	9	11	8	2	1	
64—65	1	5	7	10	3			
62—63	1	2	7	1	2			
60—61	1	1		1				

Ans. $r = \cdot 60$
$PE_r = \cdot 04$

2. Calculate r from the following table and also use both the rank difference and gain methods of calculating the coefficient of correlation and compare the results.

Boys	Marks in English	Marks in Science
A	5	21
B	15	14
C	10	28
D	20	7
E	25	35
F	30	42

Ans. $r = \cdot 49$ (App.)

3. Compute the value of r using the Pearson's product moment formula. Use both the methods, when deviations are taken from the actual means and when deviations are taken from the assumed means.

X	Y
88	65
87	69
85	71
84	59
81	55
79	40
76	61
75	60
73	57
72	63

Ans. $r = \cdot 355$

4. Compute the Pearson's Product-Moment Correlation from the following table.

Class Interval	60—	70—	80—	90—	100—	110—	120—	130—
130—						3	1	
120—					1	2		
110—				1	4	1		
100—				3	12	1		
90—					13			
80—			1	1	3			
70—		1			1			
60—								
50—	1							
40—								

Ans. $r = \cdot 703$

5. Find the product-moment coefficient of correlation from the following data :

X	Y
1	2
2	5
3	3
4	8
5	7

Ans. $\sigma_x = 1{\cdot}414$; $\sigma_y = 2{\cdot}280$; $r = +0{\cdot}81$.

6. Calculate ρ and R from the following Data :

Salesmen	Year of Service	Rank in order of efficiency
A	5	6
B	2	12
C	10	1
D	8	9
E	6	8
F	4	5
G	12	2
H	2	10
I	7	3
J	5	7
K	9	4
L	3	11

Ans. $\rho = {\cdot}80$; $R = {\cdot}56$

Hints : In this question rank in order of efficiency is already given.

Find out the rank in order of service.

Then find out the difference of ranks in each case.

Using the formula $\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$ get the answer.

7. The rankings of ten students in mathematics and music are as follows :

Mathematics : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Music : 6, 5, 1, 4, 2, 7, 8, 10, 3, 9

What is the coefficient of rank correlation.

Ans. $\rho = +0{\cdot}45$

8. Ten competitors in a beauty contest are ranked by three Judges in the following orders :

1, 6, 5, 10, 3, 2, 4, 9, 7, 8
3, 5, 8, 4, 7, 10, 2, 1, 6, 9
6, 4, 9, 8, 1, 2, 3, 10, 5, 7

Use the rank correlation coefficients to discuss which pair of judges has the nearest approach to common tastes in beauty.

Ans. If the judges be denoted by 1, 2, 3, then $\rho_{12} = -0{\cdot}21$; $\rho_{23} = -0{\cdot}30$: $\rho_{13} = +0{\cdot}64$. This suggests that judges 1 and 3 have tastes in common.

9. Calculate the product moment coefficient of correlation between the class marks in science achievement test of 110 students given below.

	10—19	20—29	30—39	40—49	50—59	60—69
60—69			1	1		2
50—59		2	6	8	3	
40—49	1	6	20	4	2	
30—39	5	5	15	7		
20—29	3	12	5			
10—19			2			

Ans. $r = {\cdot}52$.

10. Compute the coefficient correlation r from the following Table

Scores	18-21	22-25	26-29	30-33	34-37	38-41	42-45	46-49	50-53	54-57	58-61	62-65	66-69	70-73	74-77	78-81	82-85	86-89	90-93	94-97
105 102																				1
101 98																	1			
97 94																				
93 90														1		2	1	1		
89 86															2	4	1	1		
85 82													4	4	7	1	1			
81 78										1	2	1	8	5	4	2				
77 74									1	1	1	8	8	6	5	1	1			
73 70								1	2	4	12	21	15	5	5					
69 66								1	5	10	33	17	8	4						
65 62							2	7	5	21	19	14	6							
61 58						2	7	12	29	17	7	7								
57 54						2	10	11	17	8	8	1								
53 50				1	4	17	10	17	13	4	2									
49 46				3	5	8	9	5	7											
45 42				3	5	9	6													
41 38				2	1	2	2	1												
37 34				3	1	2														
33 30		1		1																
29 26	1	2	1																	

Ans. $r = \cdot376$

Hints ! $\Sigma xy = 4919$ $N = 600$

CHAPTER 18

REGRESSION AND PREDICTION

How to predict one variable from another

LET US TAKE THE EXAMPLE OF THE MARKS OF 100 COLLEGE students in Hindi and English as given in the table (7). As before (shown in table 7) we can find out the means of the columns and the rows. Suppose in this group of 100 students we wish to estimate the marks of a student in English who secures 55 marks in Hindi. The best guess that we can make is to take the average marks in English of all the 24 students those fall in the interval (Hindi Mark) 50–59. The mean mark in English of these 24 students is 49·5. Thus we can estimate that a student who obtains 55 marks in Hindi secures 49·5 marks in English.

Again let us guess the marks of a student in Hindi who secures 39 marks in English. We see the mean marks in Hindi of 15 students in the interval (English Mark) 35–44 is 44·5. Naturally we can say that a student securing 39 marks in English obtains 44·5 marks in Hindi.

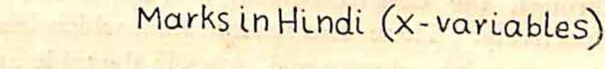
These approximate predictions are done with the help of the scatter-diagram. But in doing this much error is introduced.

While we made use of the fact that the means are the most, probable points in our columns and rows we have made no use of our knowledge concerning the correlation between the two variables. As is discussed in the previous chapter that regression lines are determined by the correlation between the two variables and the degree of their closeness denotes the coefficient of correlation, naturally we should expect that the regression lines rather than the short straight lines joining the mean will predict a better relationship between the two variables.

If we know the marks of a student in Hindi (*x*—variables) we can predict his marks in English (*y*—variables) by substituting the value of *x*-variable in the equation connecting *x* and *y*. These equations which connect *x* and *y* variables and *y* and *x* variables are called the lines of regression. Our purpose is now to develop these regression equations which may either predict the value of *x* knowing that of *y* or predict the value of *y* knowing that of *x*.

[Data from Table 7]

An illustration of the positions of Regression lines and calculations of the regression equations. (*marks in Hindi and English of 100 college students*)



	20-29	30-39	40-49	50-59	60-69	f(y)	Row Mean
55-64					6	6	64.5
45-54				24	1	25	54.9
35-44			15			15	44.5
25-34		8	29			37	42.3
15-24	5	12				17	32.1
f(x)	5	20	44	24	7	100	
Col. Mean	19.5	23.5	32.9	49.5	58.07		

$r = \cdot918$

Calculation of regression Equations.

(1) Deviation Form.

(i) $\bar{y} = \cdot918 \times \frac{11\cdot93}{9\cdot56}x = 1\cdot14x$

(ii) $\bar{x} = \cdot918 \times \frac{9\cdot56}{11\cdot93}\, y = \cdot73\, y$

(2) Score Form

(i) $\overline{y} = 1\cdot14x - 15\cdot54$

(ii) $\overline{x} = \cdot73y + 18\cdot94$

$M_{(y)} = 36\cdot1$

$M_{(x)} = 45\cdot3.$

Regression Equations

In table **7** we have already seen that by joining the row means 64·5, 54·9, 44·5, 42·3, 32·1 by short lines and the column means 19·5, 23·5, 32·9, 49·5, 58·07 by another short line we get two sets of short lines which are jagged in appearance. Two straight lines have been drawn in to describe the general trend of these irregular lines. These two lines pass as closely as possible through the dots (row-mean) and the crosses (column-mean) respectively. These two straight lines which best fit the means of the successive columns and rows in the table are called the lines of regression.

The equation of the regression line through the crosses is

$$\overline{\ } = r\ \frac{\sigma_y}{\sigma x} \times x$$

This is the regression equation of y on x where deviations are taken from the means of y and x. In this equation y denotes that the estimate is an average value, r is the coefficient of correlation between the two variables. σ_x and σ_y are respectively the standard deviations of x and y variables.

$r\ \frac{\sigma_y}{\sigma_x}$ is called the regression coefficient.

Regression coefficient can otherwise be written as b_{yx} or b_{12}.

In short the above regression equation can be written as

$\overline{y} = b_{xy} \times x$ or $\overline{y} = b_{12} \times x$.

As calculated before

$r = \cdot918$

$\sigma_x = 9\cdot56$

$\sigma_y = 11\cdot93$

Substituting these values in the above regression equation

$$\overline{y} = \cdot 918 \times \frac{11\cdot 93}{9\cdot 56} x$$

or $y = 1\cdot 14\ x$

This equation is helpful in predicting the relationship between the deviations from the mean of the English marks and the deviations from the mean of the Hindi marks

Suppose $x = 1$; then from the above relation $\overline{y} = 1\cdot 14$.

This means a deviation of one mark from the mean of Hindi marks corresponds to a deviation of 1·14 marks from the mean of English marks. Let us find the marks in English of a student who secures 55 marks in Hindi. 55 is 9·7 above the mean of Hindi marks of the group because the mean value of Hindi marks is 45·3.

Deviation of 1 mark from the mean of Hindi marks corresponds to a deviation of 1·14 marks from the mean of the English marks.

Therefore deviation of 9·7 marks from the Hindi marks will correspond to a deviation of $9\cdot 7 \times 1\cdot 14$ marks i.e. 11·05 marks from the mean of English marks.

Thus a student who secures 55 marks in Hindi will obtain $(36\cdot 1 + 11\cdot 05)$ or 47·15 marks in English.

The regression equation $\overline{y} = r\,\frac{\sigma_y}{\sigma_x} \times x$ gives the relationship between y and x in deviation form.

The regression equation of x on y, deviations taken from the mean of x and y is given by

$$\overline{x} = r\frac{\sigma_x}{\sigma_y} \times y$$

Here $r\,\frac{\sigma_x}{\sigma_y}$ is called the regression coefficient and can be otherwise written as b_{xy} or b_{21}.

So the above formula can be written in short as

$$\overline{x} = b_{xy} \times y \text{ or } \overline{x} = b_{21} \times y$$

Again substituting the value of r, σ_x and σ_y

$\overline{x} = \cdot 918 \times \frac{9 \cdot 56}{11 \cdot 93} y$

or $\overline{x} = \cdot 73\, y$

Suppose $y = 1$

Then $\overline{x} = \cdot 73$.

This shows that a deviation of one mark from the mean of the English scores corresponds to a deviation of ·73 marks from the mean of the Hindi scores.

Let us find the marks of a boy in Hindi who secures 39 marks in English.

39 is 2·9 above the mean of the English marks of the group because the mean marks of English is 36·1.

Since the deviation of one mark from the mean of the English marks corresponds to the deviation of ·73 marks from the mean marks in Hindi.

Therefore the deviation of 2·9 marks from the mean marks of English will correspond to a deviation of $\cdot 73 \times 2 \cdot 9$ or 2·11 marks from the mean marks of Hindi.

Thus a boy who secures 39 marks in English will obtain $(45 \cdot 3 + 2 \cdot 11)$ or 47·41 marks in Hindi.

Equation $\overline{x} = r \frac{\sigma_x}{\sigma_y} \times y$ gives the relationship between x and y in deviation form. It is to be noted that neither of the above regression lines can be used to predict the value of x and y simultaneously. The regression line $\overline{y} = r \frac{\sigma_y}{\sigma_x} \times x$ can be used to predict the value of y when the value of x is known.

The regression line $\overline{x} = r \frac{\sigma_x \times y}{{}_s\sigma_y}$ can be used to predict the value of x when value of y is known. When the value of r, i.e., coefficient of correlation is $+1$ or -1 these regression lines coincide.

This can be illustrated as below :

Let $r = +1$

Then $\overline{y} = r\frac{\sigma_y}{\sigma_x}x$ becames $\overline{y} = \frac{\sigma_y}{\sigma_x}x$

$$\text{or}\quad \overline{y}\ \sigma_x = \sigma_y x \qquad \text{(A)}$$

Again $\overline{x} = \frac{\sigma_x}{\sigma y}y$ becomes $\overline{x} = \frac{\sigma_x}{\sigma_y}y$

$$\text{or}\quad \overline{x}\ \sigma_y = y\sigma_x \qquad \text{(B)}$$

Equations (A) and (B) are identical and hence the two regression lines coincide. Taking a numerical example, suppose $r=1$, $\sigma_x = 9{\cdot}56$ and $\sigma y = 11{\cdot}93$

Then $\overline{y} = r\frac{\sigma_y}{\sigma_x}x$ becomes $y = \frac{11{\cdot}93}{9{\cdot}56}x$

$$\text{or}\quad 9{\cdot}56\overline{y} = 11{\cdot}93\,x \quad \text{(C)}$$

Again $\overline{x} = r\frac{\sigma_x}{\sigma_y}y$ becomes $\overline{x} = \frac{9{\cdot}56}{11{\cdot}93}y$

$$\text{or}\quad 11{\cdot}93\,\overline{x} = 9{\cdot}56\,y \quad \text{(D)}$$

Equation (C) and (D) are similar and hence both the regression lines coincide.

Similarly it can be shown that when $r = -1$ both the lines of regression coincide.

The Regression equations in scores form

The regression equations $\overline{y} = r\frac{\sigma_y}{\sigma_x}x$ and $\overline{x} = r\frac{\sigma_x}{\sigma_y}$ are in deviation forms. Now let us discuss the regression equations in scores forms. These equations in score forms are more convenient in predicting the value of one variable knowing the value of other variable. Suppose in table (7) Mx is the mean of the Hindi scores and X to be any general score in Hindi. The deviation of this X score from its mean can be written as (X$-$Mx) and let it be denoted by x.

Similarly let M y be the mean of the English scores and Y be any general score in English. The deviation of this Y score from its mean can be written as (y $-$ My) and let it be denoted by y.

Substituting these values of x and y in the regression equation ; $\overline{y} = r\frac{\sigma_x}{\sigma_y}x$ and $\overline{x} = r\frac{\sigma_x}{\sigma_y}y$;

We get the two regression equations in the score forms i.e.

$$\overline{Y}-M_y = r\,\frac{\sigma_y}{\sigma_x}\,(X-M_x) \text{ or } \overline{Y}=r\,\frac{\sigma_y}{\sigma_x}\,(X-M_x)+M_y$$

$$\text{and } \overline{X}-M_x=r\,\frac{\sigma_x}{\sigma_y}\,(Y-M_y) \text{ or } \overline{X}=r\,\frac{\sigma_x}{\sigma_y}\,(Y-M_y)+M_x$$

Here X and Y in both the equations represent the actual scores and not deviations from the means of the two distribution.

On substituting the values of M_y, r, σ_x, σ_y and M_x in the above equations we get

$$\overline{Y}=\cdot918\times\frac{11\cdot93}{9\cdot56}\,(X-45\cdot3)+36\cdot1$$

$$\text{or} \quad \overline{Y}=1\cdot14\,X-15\cdot54 \qquad \text{(E)}$$

Which is regression equation of the English marks on the Hindi marks in the score form.

$$\overline{X}=\cdot918\times\frac{9\cdot56}{11\cdot93}\,(Y-36\cdot1)+45\cdot3$$

$$\text{or} \quad \overline{X}=\cdot73\,Y+18\cdot94 \qquad \text{(F)}$$

This is the regression equation of the Hindi marks on the English marks in the score form.

Let us predict the marks obtained by a student in English who secures 55 marks in Hindi.

Substituting the value of X as 55 in equation (E)

We get $\overline{Y}=1\cdot14\times55-15\cdot54$

or $\overline{Y}=47\cdot16$

Thus a student who secures 55 marks in Hindi obtains 47·16 marks in English. The same result we get from the regression equation in the deviation form.

Again let us predict the marks obtained by a student in Hindi who secures 39 marks in English.

Substituting the value of Y as 39 in equation (F)

We get $\overline{X} = \cdot 73 \times 39 + 18\cdot 94$

or $\overline{X} = 47\cdot 41$

Thus a student who secures 39 marks in English obtains 47·41 marks in Hindi. The same result we got from the regression equation in deviation form.

SUMMARY

(1) $\bar{y} = r\,\frac{\sigma_y}{\sigma_x} \times x$

This is the regression equation of y on x, deviations being taken from the means of Y and X. The factor $r\frac{\sigma_y}{\sigma_x}$ is the regression coefficient.

(2) $\bar{x} = r\,\frac{\sigma_x}{\sigma_y} \times y$

This is the regression equation of x on y, deviations being taken from the means of X and Y. The factor $r\frac{\sigma_x}{\sigma_y}$ is the regression coefficient.

(3) $\overline{Y} = r\,\frac{\sigma_y}{\sigma_x}(X - M_x) + M_y$

This is the regression equation of Y on X in the score form.

(4) $\overline{X} = r\,\frac{\sigma_x}{\sigma_y}(y - M_y) + M_x$

This is the regression equation of X on Y in the score form.

EXERCISE

1. Given the following data for two tests :

History (X)	English (Y)
$M_x = 75\cdot 00$	$M_y = 70\cdot 00$
$\sigma_x = 6\cdot 00$	$\sigma_y = 8\cdot 00$

$r_{xy} = \cdot 72$

Work out the regression equation in score form.

Ans. $\overline{Y} = \cdot 96\,X - 2$; $\overline{X} = \cdot 54\,Y + 37\cdot 2$

2. Write out the regression equations in score form for the correlation table given below—Boys Ages 4·5 to 5·5 years

Weight in pounds (X)

Height in inches (Y)	24—28	29—33	34—38	39—43	44—48	49—53	Totals
45—47			1		2		3
42—44			4	35	21	5	65
39—41		5	87	90	7	1	190
36—38	1	18	72	8			99
33—35	5	15	5				25
30—32	2						2
Totals	8	38	169	133	30	6	384

Ans. $\bar{Y} = \cdot 40X + 24\cdot 12$

$\bar{X} = 1\cdot 26Y - 11\cdot 52.$

3. Given the following data for 56 children ;

X_1 = Stanford – Binet I. Q.
X_2 = Memory for objects.
X_3 = Cube imitation

$M_1 = 101\cdot 71$	$M_2 = 10\ 06$	$M_3 = 3\ \cdot 35$
$\sigma_1 = 13\cdot 65$	$\sigma_2 = 3\cdot 06$	$\sigma_3 = 2\cdot 02$
$r_{12} = 0\cdot 41$	$r_{13} = 0\cdot 50$	$r_{23} = 0\cdot 16$

Work out the regression equation of X_2 and X_3 upon X_1.

4. Work out the regression equation both in the deviation form and the score form for the following data :

Weight (X)
$M_x = 136{\cdot}3$
$\sigma_x = 15{\cdot}55$

Height (Y)
$M_y = 66{\cdot}5$
$\sigma_y = 2{\cdot}62$

$r_{xy} = {\cdot}60$

Answer.—

Regression equation in deviation form

(1) $\bar{y} = {\cdot}10x$
(2) $\bar{x} = 3{\cdot}56y$

Regression equation in score form

(1) $\bar{Y} = {\cdot}10X + 52{\cdot}9$
(2) $\bar{X} = 3{\cdot}56Y - 100{\cdot}4$.

CHAPTER 19

GRAPHIC METHODS

Why Graphic Methods

IN STATISTICAL METHOD WE GENERALLY USE THE DATA IN the tabular form. This data in the tabular form reveals different kinds of facts to the reader. But if this statistical data is put in the form of a graph it gives the lasting effect and makes it easy for one to compare the different data depicted by the graphs. For example suppose we want to compare the science achievement of the eleven year old pupils of India and England. For this we shall take a sample of eleven year old pupils from India and another sample of eleven year old pupils from England. After applying some sort of test and collecting the scores of the samples we shall draw two graphs for the two sets of scores. These graphs shall help us to compare the science achievement of the eleven year old pupils of the two countries.

Graphical representation of the data has another advantage too. We all know that the visual presentation is far more effective than complex numerical presentation and for this reason graphical method is preferred.

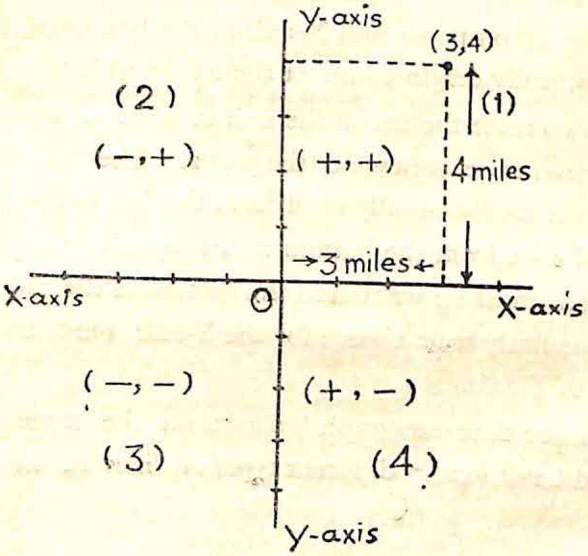
In the schools the teachers can make use of this method in plotting the progress of the pupils in various school subjects each month. It may be used for comparing the achievement of the pupils of different sections of a class.

Graphs (General Principles)

The graphs are drawn on a graph paper. A vertical line on it represents the Y-axis and a horizontal line represents the X-axis. The meeting point of the Y-axis and the X-axis is called the origin which is represented by the letter O. The X-axis and the Y-axis divides the graph paper into four quadrants. The following figure shows the four quadrants.

On each of the X-axis and the Y-axis we take certain equal distances which represent the measure of the behaviour in terms of the unit of the distance on these axes. It is not necessary that we should take the same unit in the X-axis and the Y-axis.

All the points above X-axis are positive and below it are negative. All the points on the right of X-axis are positive and on the left negative. We see that all the co-ordinates in quadrant (1) are (+, +); in (2) (—, +); in (3) (—, —) and in (4) (+, —). If we ask a man to go three miles to the east of the origin O on the X-axis and then 4 miles north he will reach a point whose co-ordinates are called (3, 4). Thus if each division on the X-axis and Y-axis represents one mile he will walk 3 division on the X-axis and this distance is [called abscissa'. Similarly he will walk four divisions up and this distance along the Y-axis is called the 'ordinate'. The point (3, 4) lies in quadrant (1).



Again if we ask the man to walk 3 miles west of the origin on the X-axis and then 4 miles north of it, the co-ordinates of the point where he will reach will be (−3, 4) and this will lie in quadrant (2). Similarly point (—3, —4) will lie in quadrant (3) and (+3, -- 4) will lie in quadrant (4). It is not essential that if one division represents one mile on the X-axis, the division of the same magnitude on the Y-axis should also represent one mile. We can assume that 2 divisions on the Y-axis represent one mile. We can choose the units for abscissa and ordinate according to our convenience but it must be so chosen as to give a good shape to the graph. If we are given the co-ordinate of different points we can plot them on the graph paper. Then these points are joined which represents the graph for these points. These graph lines should be sharply defined so that they may be distinguished from the other rulings of the graph paper.

In Statistical calculations we generally require the quadrant (1) and in this quadrant all our co-ordinate of statistical data lie. Thus the graph covers this quadrant only. To avoid the misuse of the paper we generally take the Y-axis on the left corner of the graph paper and X-axis at the bottom of the graph paper. Naturally origin comes at the left hand bottom corner.

We can increase the size of the unit to give a good shaped graph. It also saves wastage of the paper. The horizontal scale for the curves should usually read from the left to the right and the vertical scale from the bottom to the top. The scale chosen on the X-axis must be written at the bottom of the paper below X-axis. Similarly scale chosen for the Y-axis must be written on the left of Y-axis.

When more than one graph is shown on the same diagram they should be drawn in different types of lines or in different colours.

Special Graphs

The following types of the graphs are generally used in statistics.

(a) The frequency polygon.

(b) The Histogram.

(c) The cumulative frequency curve.

(d) The cumulative percentage curve.

The Frequency Polygon

Suppose we have to draw the graph for the frequency distribution given below :

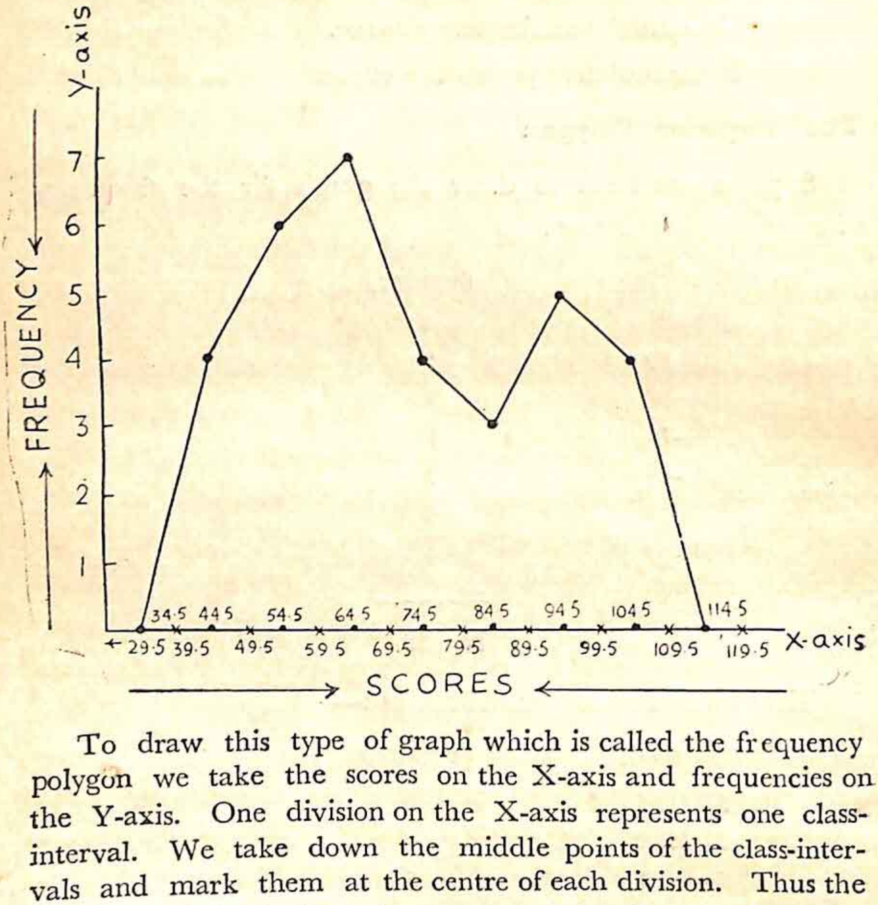
Class-intervals	frequency
100—109	4
90—99	5
80—89	3
70—79	4
60—69	7
50—59	6
40—49	4
	N=33

The distribution can be written as :—

Class-intervals	Frequency
109 5—119·5	0
99·5—109·5	4
89·5— 99·5	5
79·5— 89 5	3
69·5— 79·5	4
59·5— 69·5	7
49·5— 59·5	6
39 5— 49·5	4
29·5— 39·5	0

GRAPH No. 1.

Frequency Polygon



To draw this type of graph which is called the frequency polygon we take the scores on the X-axis and frequencies on the Y-axis. One division on the X-axis represents one class-interval. We take down the middle points of the class-intervals and mark them at the centre of each division. Thus the class-interval 39·5—49·5 is represented by the second division and 44·5 is its middle value. The scores between 39·5 – 49·5 are represented by 44·5 on the X-axis. The frequencies are taken on the Y-axis. Here each division on the Y-axis represents one frequency. Since in the class-interval 39·5—49·5 whose middle value is 44·5, the number of frequencies are 4, we move four divisions up from 44·4 which is on the X-axis and plot that point. Similarly mid-value of 59·5—69·5 is 64·5. Frequencies in this interval are 7. Hence we move 7 divisions up from 64·5. Similarly other points are plotted and joined as shown in the graph. Here one thing is to be noticed. The lowest given class-interval is 39·5—49·5 and highest one is 99·5—109·5. But we take one more class-interval below the lowest one, i.e., 29·5—39·5. We also take one class-interval more above the highest one, i.e., 109·5—119·5. It is supposed that frequencies in these intervals will be zero in each case.

Mid.-value of 29·5—39·5 is 34·5 and in this interval frequency is zero. So we plot a point at 34·5 on the X-axis. Similarly a point at 114·5 on the X-axis denotes the zero frequency in the class-interval of which 114·5 is the middle point. We take these two points in our graph to close the surface represented by the graph. Thus joining all the points we get the frequency polygon as shown in the diagram above. The total number of frequencies (N) of a distribution is represented by the area of the polygon ; that is the area bounded by the frequency surface and the X-axis.

We must choose such a scale on the X-axis and the Y-axis which should give a good look to the figure. Generally we select X and Y-units, such that the height of the figure is approximately 75% of its width. Thus in the present figure we have 8 class-intervals on the X-axis including two half class-intervals in the beginning and the end. These 8 class-intervals cover 8 divisions. The highest frequency is 7, in the class-intervals 59·5—69·5. We now take the height of our figure as

$$\frac{75\times 8}{100}=6 \text{ X-units.}$$

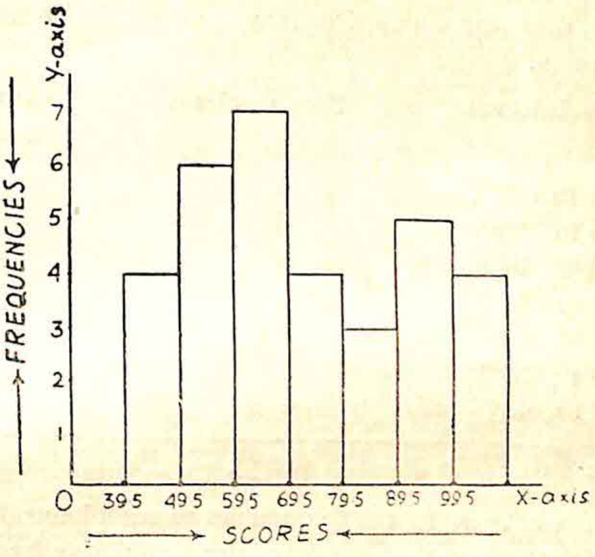
Thus we can take the scale on the Y-axis as one division representing one frequency because highest frequency is 7 and it should roughly lie within 6 X-units height to give the accurate figure. It means that almost 1 division is equal to 1 frequency on the Y-axis. It will not affect our figure if we take one division on the Y-axis to represent one frequency. Thus 7 divisions on the Y-axis represents 7 frequencies. We can change the ratio of the height of the figure to its width from 60% to 80%. We must note here that the area lying above any given interval cannot be taken as proportional to the number of cases within the interval because of the irregularities in the distribution and consequently in the frequency surface.

The Histogram.

There is another way of representing the frequency distribution by means of a graph called the histogram. Here we choose the scale on the X-axis and the Y-axis in the same way as is done in the case of the frequency polygon. Scores are taken on the X-axis and frequencies on the Y-axis. The ratio of the height of the figure to its width is taken almost as 75% but it does not affect our figure if we take this ratio varying from 60% to 80%. Suppose we have to draw a histogram for the same data which we used for drawing a frequency polygon. In this case we suppose one class-interval is represented by one division on the X-axis and one division on the Y-axis represents one frequency. The frequencies in the class-interval 39·5—49·5 are 4. So we erect a rectangle with its base as this class-interval (39·5—49·5) and height as 4 units of the Y-axis. Similarly frequencies in the class-interval 59·5—69·5 are 7. Hence we erect a rectangle having its base the class-interval 59·5—69·5 and height as seven units of the Y-axis. In this way rectangles are erected on each class-interval. This figure is called the histogram. As in the case of the frequency—polygon, here as well the total number of frequency (N) is given by the whole area of the histogram, i.e., the sum of the area of all the ractangles erected. In the case of histogram the area represented by each rectangle is proportional to the number of measures within the interval. In this respect it differs from the frequency polygon. Histogram presents us the picture of the relative proportions of the total frequency from interval to interval.

THE HISTOGRAM

Data from Graph No. 1



Graph No. 2

Comparative Study of the Histogram and the Frequency Polygon

We have seen that both the types of graphs give the picturesque view of the statistical data. As discussed before the histogram is more exact than the frequency polygon because in the case of the frequency polygon we have the irregularities in the frequency surface and for which reason it fails to represent the number of measures in each interval. The histogram does so very accurately. But when we are to compare two or more graphs in the same diagram with the same axis we always prefer the frequency polygon because the horizontal and vertical lines of the histogram often coincide and arouses confusion.

The cumulative frequency graph. The cumulative frequency graph is another way of graphical representation of the frequency distribution. If we are given the frequencies in the different intervals, we can change the table to one, containing cumulative frequencies in the same way as discussed in chapter 3 under head 'Cumulative Frequency'.

Suppose we have to draw the cumulative frequency graph for the data for which frequency polygon and histogram have been drawn.

Class-Interval	Frequencies (f)	Cumulative frequencies
99·5 to 109·5	4	33
89·5 to 99·5	5	29
79·5 to 89·5	3	24
69·5 to 79 5	4	21
59·5 to 69·5	7	17
49·5 to 59·5	6	10
39·5 to 49·5	4	4

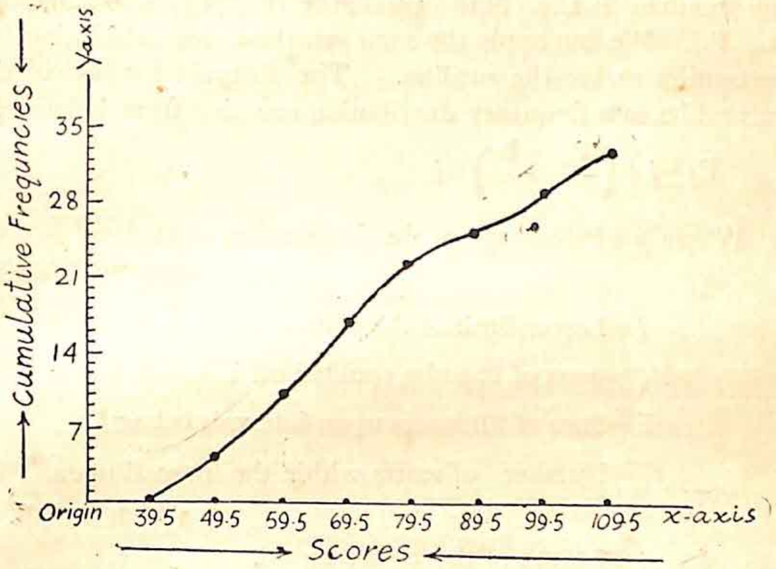
N = 33

First we convert the frequencies to cumulative frequencies. Then for plotting the graph of the cumulative frequencies we take the X-axis as presenting the scores and the Y-axis as the cumulative frequencies. In the case of the frequency polygon we took the middle point of the class-interval but in this case we take the upper limit of each class interval which are 49·5, 59·5,...109·5. In the class interval 39·5—49·5 the cumulative frequency is 4, so we move upto 49·5 on the X-axis and then 4 units up above the Y-axis and plot the point. Similarly in the next interval we have 10 cumulative frequencies. We plot a point 10 units up of Y-axis above the point 59·5 and so on. Thus we get a 'S' shaped graph which is called the cumulative frequency graph.

In this case we have 7 intervals and by 75% rule there should be $\frac{75}{100} \times 7 = 5$ units distance (each equal to one class-interval) on the Y-axis. Since the highest cumulative frequency is 33, each of these Y units should represent $\frac{33}{5}$ or approximately 7 scores. We divide the Y-axis into 5 equal parts and each part repsesents 7 cumulative frequencies.

The Cumulative Frequency Graph

[Data from the table of frequency polygon.]



Graph No. 3

In order to begin the curve from the X-axis we take one class-interval more below the lowest interval. Naturally its upper limit will be 39·5. Since in this class-interval cumulative frequency will be zero, hence the point will be on the X-axis at 39·5.

Percentile and Percentile Ranks

Calculation of percentiles in a frequency distribution. We know that median is a point in the distribution above and below which lie 50% scores. In the case of first quartile Q_1 and third quartile Q_3 we have read that these are respectively the points in the distribution below which lie 25% and 75% cases. We called median as the second quartile i.e. Q_2. The methods of computing median, first quartile and third quartile have already been discussed. If we are to compute points below which lie 20%, 10%, 99%, 95% cases etc. we use the same principle as is used in the case of first and third quartiles.

These points are called percentiles and are denoted by P_p where p refers to the percentage of cases below the given value. Thus P_{20} is a point below which lie 20% of the scores. Now we can say median is the 50th percentile (P_{50}), Q_1 is P_{25} and Q_3 is P_{75}. We can apply the same method for calculating the percentiles as for the median. The formula for calculating percentiles in a frequency distribution counting from below up is

$$P_p = l + \left(\frac{P_N - F}{f_p}\right) \times i$$

Where p = Percentage of the distribution wanted e.g. 20%, 40%, 50%, etc.

l = Lower limit of the class-interval in which p lies.

P_N = part of N to be counted off in order to reach P_p.

F = Sum of all scores upon intervals below l.

f_p = Number of scores within the interval upon which P_p fall.

i = Length of the class-interval.

Let us take the numerical example for calculating the percentiles with the help of the above formula. Suppose we have to calculate P_{60} for the following table.

Class interval	Frequency f	Cumulative frequency (f)
70—79	11	40
60—69	9	29
50—59	7	20
40—49	5	13
30—39	3	8
20—29	3	5
10—19	2	2

N=40

Here $P_N = \frac{60}{100} \times 40 = 24$ i.e. (60% of 40)

From the above cumulative frequency table we find that 20 scores lie upto 59 5 the lower limit of the interval 60–69 and 29 scores lie upto 69·5 the upper limit of the class-interval 60—69.

Hence P_{60} falls upon 60–69.

Thus $F=20$ (sum of the scores upon intervals below l)

$P_N=24$

$f_p=9$ (number of scores within the interval upon which P_{60} falls)

$l=59{\cdot}5$

$i=10$

$$\therefore\ P_{60}=5{\cdot}95+\frac{(24-20)}{9}\times 10$$

$$=59{\cdot}5+\frac{4}{9}\times 10$$

$$=63{\cdot}9=64 \text{ (approximately)}$$

This means 60% of the scores lie below 64 in the distribution. Thus the value of P_{60} is 64.

The other percentile values can be calculated in the same way as P_{60}. Below are given a few percentile values for this very distribution.

It should be noted that P_0, which marks the lower limit of the first interval (namely 9·5) lies at the beginning of the distribution. P_{100} marks the upper limit of the last interval i.e. 79·5 and lies at the end of the distribution. These two percentiles represent limiting points.

Calculation of percentiles (Decile points)

70% of 40 = 28	$59{\cdot}5+\frac{(28-20)}{9}\times 10=68{\cdot}3$
60% of 40 = 24	$59{\cdot}5+\frac{(24-20)}{9}\times 10=63{\cdot}9$

50% of 40 = 20	$59{\cdot}5+\frac{(20-20)}{7}\times 10=59{\cdot}5$
40% of 40 = 16	$49{\cdot}5+\frac{(16-13)}{7}\times 10=53{\cdot}8$
30% of 40 = 12	$39{\cdot}5+\frac{(12-8)}{5}\times 10=47{\cdot}5$
20% of 40 = 8	$39{\cdot}5+\frac{(8-8)}{3}\times 10=39{\cdot}5$
10% of 40 = 4	$19{\cdot}5+\frac{(4-2)}{3}\times 10=26{\cdot}1$

Calculation of Percentile Ranks in a frequency Distribution

Uptil now we have read how percentiles e.g. P_{60}, P_{70} etc. are calculated from a frequency distribution. We called percentiles as points in a frequency distribution below which lie a given percentage of total scores (N). In the case of percentile rank we find a position on a scale of 100 to which the subject's score entitles him. The method of calculating the percentile rank is reverse of the method of calculating percentile.

Let us illustrate this with the same table which we used for percentiles. Let us find the percentile rank (PR) of a man who scores 65. This score 65 falls in the interval 60—69. There are 20 scores upto 59·5, lower limit of this class interval and 9 scores spread over this interval.

In an interval of 10 units lies 9 scores, therefore in an interval of one unit lies $\frac{9}{10}=\cdot 9$ scores. The score of 65 is 5·5 score units from 59·5 (lower limit of the interval within which the score of 65 lies).

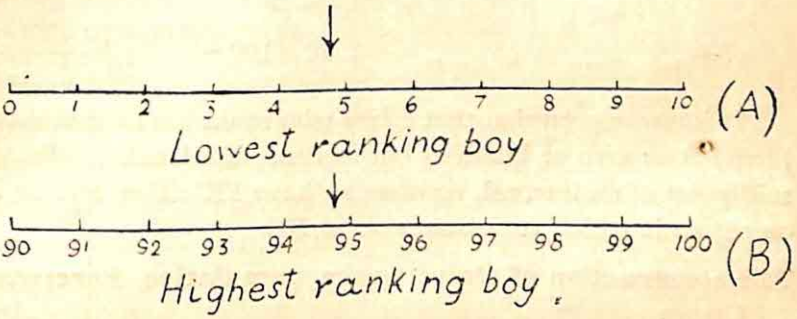
5·5 score units $=5{\cdot}5\times\cdot 9$ score distance of 65 from 59·5
$=4{\cdot}95$ score distance of 65 from 59·5

Adding 4·95 to 20 (number of scores below 59·5) we get 24·95 as the part of N lying below 65. Dividing 24·95 by 40 and multiplying it by 100 we get 62·3% as the proportion of N below 65. Hence the percentile rank of score 65 is 62.

Calculation of percentile Ranks when individuals or objects are inorder of merit.

Percentile ranks are generally used in Experimental Psychology when we are dealing with attributes for which individuals or objects are to be arranged in order of merit but in which they cannot be measured directly.

Let us illustrate the use of PR's by taking an example in which 10 boys are to be ranked as 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 in order of merit for aesthetic qualities. Here the highest-ranking boy has a percentile 95 and the lowest ranking boy has a percentile 5, as shown below. On scale running from 0—100 each of 10 individuals occupies 10 divisions $\left(\frac{100}{10} \text{ or } 10\%\right)$. The last ranked boy is assigned the mid-point of the first ten divisions of the scale (0—10) which is 5 and the first ranked boy is assigned the mid-point of the highest 10 divisions which is 95. The following diagrams (Diagram A and B) illustrate the assigning of percentile ranks to the best and the poorest boy in a group of 10 boys.



The following formula is used in converting orders of merit into equivalent percentile Ranks.

$$PR = 100 - \frac{(100R - 50)}{N}$$

Applying this formula for calculating the PR of the boys who rank 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 respectively we get the following results.

(1) PR of a boy whose rank is one.

$$= 100 - \frac{(100 \times 1 - 50)}{10} = 95$$

(2) PR of a boy whose rank is $2=100-\frac{(100\times 2-50)}{10}=85$

(3) ,, ,, ,, ,, ,, $3=100-\frac{(100\times 3-50)}{10}=75$

(4) ,, ,, ,, ,, ,, $4=100-\frac{(100\times 4-50)}{10}=65$

(5) ,, ,, ,, ,, ,, $5=100-\frac{(100\times 5-50)}{10}=55$

(6) ,, ,, ,, ,, ,, $6=100-\frac{(100\times 6-50)}{10}=45$

(7) ,, ,, ,, ,, ,, $7=100-\frac{(100\times 7-50)}{10}=35$

(8) ,, ,, ,, ,, ,, $8=100-\frac{(100\times 8-50)}{10}=25$

(9) ,, ,, ,, ,, ,, $9=100-\frac{(100\times 9-50)}{10}=15$

(10) ,, ,, ,, ,, ,, $10=100-\frac{(100\times 10-50)}{10}=5$

It is our expectation that a boy who ranks last or first should have PR as zero or hundred but as Percentile Rank is always a mid-point of an interval, no one can have PR either zero or 100 on the scale which runs from zero to 100.

The construction of Ogive or the cumulative Percentage Curve

The cumulative percentage curve or Ogive differs from the cumulative frequency graph. In Ogive frequencies are expressed as cumulative percents of N on the Y-axis instead of cumulative frequency. If we consider the same distribution as used in the case of cumulative frequency graph, we can convert each cumulative frequency into cumulative percentage. The simple way of doing this is to multiply each cumulative frequency by means of $\frac{100}{N}$, where N is the total number of frequencies, Say if in the following table we are to convert 29 to its cumulative percentage, we multiply 29 by $\frac{100}{N}$.

The cumulative percentage of 29

$$= 29 \times \frac{100}{33} = 87{\cdot}8$$

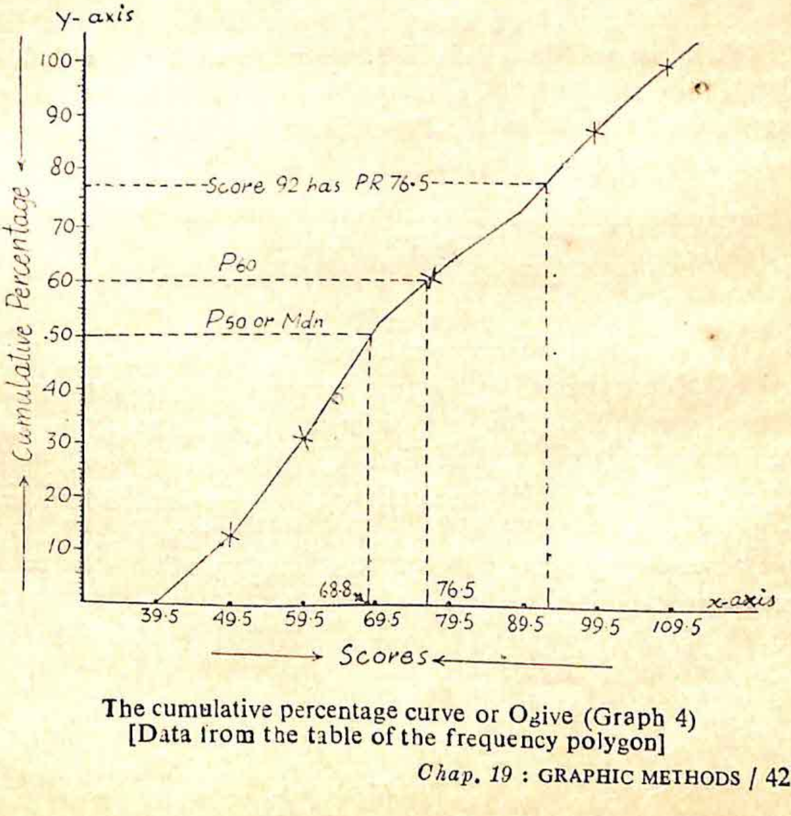
Similarly all cumulative frequencies can be converted into cumulative percents of N.

(1) Class-interval	(2) f	(3) cum. f	(4) cum. percent f
99·5 to 109·5	4	33	100·0
89·5 to 99·5	5	29	87·8
79·5 to 89·5	3	24	72·7
69·5 to 79·5	4	21	63·6
59·5 to 69·5	7	17	51·5
49·5 to 59·5	6	10	30·3
39·5 to 49·5	4	4	12·1

N=33

Thus in the above table the cumulative frequencies are expressed as cumulative percentages of N in the column (4).

To draw the Ogive with this data we proceed on the same lines as in the case of cumulative frequency graph but instead of taking cumulative frequencies on the Y-axis we take the cumulative percents of N on it.



The cumulative percentage curve or Ogive (Graph 4)
[Data from the table of the frequency polygon]

Class-intervals have been laid on the X-axis. In order to begin the curve from the X-axis we take one class-interval more below the lowest one. Naturally its upper limit will be 39·5. In this class-interval frequency is zero and therefore, cumulative frequency will be zero and also cumulative percentage will be zero.

The second point on the graph is placed at 12·1 Y-units just above 49·5 ; third point is 30·3 Y-units just above 59·5, etc. The last point is 100 Y-units above 109·5, the upper limit of the highest class-interval.

The ogive may be used to determine the percentile ranks.

As in the above case to obtain P_{50}, the median, we draw a line from 50 on the Y-scale parallel to the X-axis and from the point where this line cuts the curve we drop a perpendicular to the X-axis. This operation locates the median at 68·8 approximately. The exact median can be calculated by using the formula as below :

$$P_{Mdn} = \overset{5}{\cancel{4}}9\cdot5 + \frac{(16\cdot5 - 10)}{7} \times 10$$
$$= 68\cdot8.$$

Similarly we see P_{60} is 76·5. For this we begin with 60 on the Y-axis, go horizontally over to the curve, and drop a perpendicular to locate P_{60} approximately at 76·5. This can be verified by using the formula

$$P_{60} = 69\cdot5 + \frac{(19 - 17)}{4} \times 10$$
$$= 76\cdot5.$$

We can read the percentile rank of a given score from the ogive. The process will be just reverse of that applied in the case of percentiles. Say we want to read the PR of the score 92 from the ogive. Calculation consists in starting with score 92 on the X-axis, going vertically up to the ogive, and horizontally across to the Y-axis to locate the PR at 76·5 on the cumulative percentage scale. This can be verified by using the actual process as explained earlier.

The ogive can be used for comparing the achievements of two or more groups when their scores in a given test are plotted upon the same coordinate axis.

If the ogive of one group lies to the right of the other over the entire range, we can conclude that the scores of the former are consistently higher than those of the latter. The distances separating the two ogives at various levels indicate the differences in the achievement of the two groups.

This can be illustrated with an example. The author applied a general science test to 185 boys and 127 girls of the Fourth Form of the secondary schools in England. The following are the two separate frequency distribution of the scores of boys and girls.

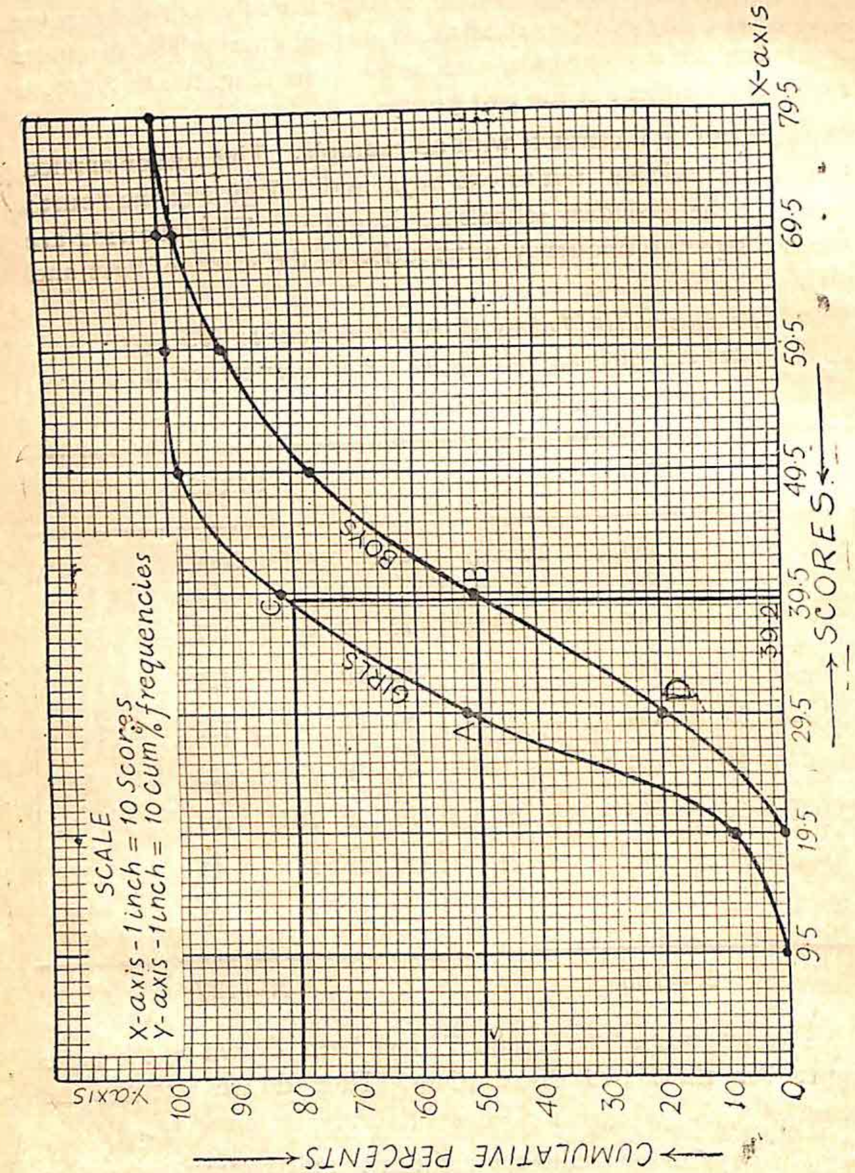
Frequency distributions of the scores, made by 185 boys and 127 girls respectively, on the achievement test in General Science.

BOYS				GIRLS			
C.I.	*f*	Cum. *f*	Cum.% *f*	C.I.	*f*	Cum. *f*	Cum. % *f*
70—79	1	185	100	70—79	0	127	100
60—69	10	184	99·3	60—69	0	127	100
50—59	28	174	93·9	50—59	4	127	100
40—49	51	146	78·8	40—49	20	123	98·4
30—39	56	95	51·3	30—39	38	103	82·4
20—29	39	39	21	20—29	53	65	52
10—19	0	0	0	10—19	12	12	9·6
N=185				N=127			

Two separate ogives (graph 5) have been drawn on the same paper for these two distributions. How do we interpret the results ?

The boys' ogive lies to the right of the girls' over the entire range, showing that the boys' score is consistently higher than the

Ogives representing the Scores made by 185 boys and 127 girls of the fourth form in the General Science test respectively



GRAPH 5.

girls. It is also clear that differences at the extremes—between the very high-scoring and the very low-scoring boys and girls—are not so great as are differences over the middle range. The boys' median is approximately 39·2, the girls' 29·5 and the difference between these measures is represented by the line AB. By extending the vertical line through B (the boys' median) upto the ogive of the girls' scores, it is clear that approximately 81% of the girls fall below the boys' median. Hence approximately 19% of the girls exceed the median of the boys' in general science test. The vertical line through A (girls' median) cuts the boys' ogive at approximately the 20th percentile. Therefore 20 p.c. of the boys fall below the girls' median, and 80 p.c. above this point. This illustrates how ogive curve is helpful for the researcher in comparing the performance of two groups.

SUMMARY

(1) **Specific Graphs—**

(1) The frequency polygon.
(2) The histogram.
(3) The cumulative frequency curve.
(4) The cumulative percentage curve or ogive.

(2) Formula for calculating the percentile is :

$$P_p = l + \left(\frac{P_N - F}{f_p} \right) \times i$$

Where : p = Percentage of the distribution wanted.
l = Lower limit of the class-interval in which p lies.
P_N = Part of N to be counted off in order to reach P_p.
F = Sum of all scores in intervals below l.
f_p = Number of scores within the interval in which P_p falls.
i = Length of the class-interval.

(3) Formula for converting orders of merit into equivalent percentile ranks.

$$PR = 100 - \frac{(100R - 50)}{N}$$

Where : R = Rank.
PR = Percentile rank.
N = Number of cases.

EXERCISES

1. The following are the scores made by 125 pupils in an achievement test in general science :

104, 90, 85, 83, 81, 80, 79, 78, 78, 77, 76, 74, 74,73, 73, 73, 72, 71, 71, 70, 69, 68, 68, 68, 67, 67, 67, 66, 66, 65, 65, 65, 65, 65, 65, 65, 64, 63, 63, 63, 63, 63, 62, 62, 62, 61, 61, 61, 60, 60, 60, 60, 60, 60, 60, 59, 58, 58, 58, 58, 58, 58, 57, 57, 56, 55, 54, 54, 54, 54, 54, 53, 53, 53, 53, 52, 51, 51, 51, 50, 50, 49, 49, 49, 49, 49, 49, 49, 48, 48, 48, 48, 48, 47, 47, 47, 45, 45, 45, 44, 44, 44, 44, 43, 43, 43, 42, 42, 41, 41, 40, 38, 38, 38, 38, 38, 37, 37, 36, 36, 35, 35, 34, 28, 25.

Make a frequency distribution for these scores and draw a frequency polygon, a histogram, a cumulative frequency curve and an ogive on four separate graph papers.

2. Draw frequency polygon, a histogram, a cumulative frequency curve and an ogive for the following fifty Army Alpha scores.

C. I.		f
195—199	...	1
190—194	...	2
185—189	...	4
180—184	...	5
175—179	...	8
170—174	...	10
165—169	...	6
160—164	...	4
155—159	...	4
150—154	...	2
145—149	...	3
140—144	...	1
		N=50

3. For the frequency distribution of the above second question calculate P_{10}, P_{20}, P_{30}, P_{40}, P_{50}, P_{60}, P_{70}, P_{80} and P_{90}.

Ans. *152·0, 159·5, 165·3, 169·5, 172·0, 174·5, 177·6, 181·5,* 187·0 respectively.

4. Find the Percentile Rank of a man who scores 163 in the same table of question 2.

Ans. 26.

5. Determine the PR of score 9 and of score 12 in the following frequency distribution.

scores	*f*
14—15	3
12—13	8
10—11	15
8—9	20
6—7	10
4—5	4
	N=60

Ans.

PR of 9=50

PR of 12=84·5

6. In the following distribution of groups (A) and (B) what is the percentile rank of score 55 in Group A's distribution ? Also in these distribution a percentile of 70 in Group (A) corresponds to what percentile in Group (B) :

C. I.	Group A	Group B
79—83	6	8
74—78	7	8
69—73	8	9
64—68	10	16
59—63	12	20
54—58	15	18
49—53	23	19
44—48	16	11
39—43	10	13
34—38	12	8
29—33	6	7
24—28	3	2
	N=128	N=139

Ans :—58, 62.

CHAPTER 20

THE NORMAL PROBABILITY CURVE

Probability

BEFORE WE DISCUSS THE NORMAL PROBABILITY CURVE WE should know something about probability.

If an event can happen in 'a' ways and fails in 'b' ways and each of the ways is equally likely, the probability or the chance of its happening is $\frac{a}{a+b}$ and that of its failing is $\frac{b}{a+b}$. Suppose there are 8 prizes and 25 blanks in a lottery. Then the probability or chance that a person holding 1 ticket will win a prize is $\frac{8}{33}$ and his probability of not winning is $\frac{25}{33}$. This illustration explains the meaning of probability. Again if we toss an unbiased coin we expect either head or tail, because there are only two faces containing head or tail. So the probability of getting head is $\frac{1}{1+1}=\frac{1}{2}$ and similarly the probability of getting tail is $\frac{1}{1+1}=\frac{1}{2}$. If we have a die which has six faces, the probability of getting each face each time is $\frac{1}{6}$. In each of the above examples if we add the probabilities of happening and not happening we get one in each case and that we call a certainty.

Thus $\frac{a}{a+b}+\frac{b}{a+b}=\frac{a+b}{a+b}=1$ (first example)

$\frac{8}{33}+\frac{25}{33}=\frac{33}{33}=1$ (second example)

$$\frac{1}{2}+\frac{1}{2}=\frac{2}{2}=1 \qquad \text{(third example)}$$

$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{6}{6}=1 \text{ (fourth example)}$$

Thus we can say if p be the probability of happening of an event, the probability of its not happening is $1-p$.

Instead of saying that the chance of happening of an event is $\frac{a}{a+b}$, it is sometimes stated that the odds are '*a*' to '*b*' in favour of the event, or '*b*' to '*a*' against the event.

If we toss a coin having two sides, one of which is head (H) and another is tail (T) then as discussed before the probability of getting H is $\frac{1}{2}$ and the probability of getting T is $\frac{1}{2}$.

In statistics this is written as below.

$$(H+T)^1=\tfrac{1}{2}+\tfrac{1}{2}=1 \qquad (1)$$

Suppose we toss two coins at the same time. This time we can have four types of combinations of the sides and any one of them has equal chance of happening. The combinations of the coins are as below.

(*i*) (*ii*)	(*i*) (*ii*)	(*i*) (*ii*)	(*i*) (*ii*)
(H) (H)	(H) (T)	(T) (H)	(T) (T)
(1)	(2)	(3)	(4)

In tossing two coins at the same time, the probability of getting first pair out of the four pairs, is $\frac{1}{4}$. Similarly the probability of getting the second pair is $\frac{1}{4}$ and the same for the rest two cases. But it makes no difference whether we get the pair in the second form or the third form. We are to get only (HT) combination whether H be on the right or on the left. So we can add the probability of the second happening and the third happening which is $\frac{1}{4}+\frac{1}{4}$ or $\frac{1}{2}$.

Thus in short probability of getting $H^2=\frac{1}{4}$.
The probability of getting $2HT=\frac{1}{2}$.
The probability of getting $T^2=\frac{1}{4}$.

If we add the left hand side we get

$$(H^2 + 2HT + T^2) \text{ or } (H + T)^2 \qquad (2)$$

The importance of this $(H+T)^2$ will be discussed after a few lines.

Again suppose we toss three coins at the same time. Here we can have 8 pairs of combinations as shown below :

(1)	(2)	(3)	(4)	(5)	(6)
(*i*) (*ii*) (*iii*)	(*i*) (*ii*) (*iii*)	(*i*) (*ii*) (*iii*)	(*i*) (*ii*) (*iii*)	(*i*) (*ii*) (*iii*)	(*i*) (*ii*) (*iii*)
(H)(H)(H)	(H)(H)(T)	(H)(T)(H)	(T)(H)(H)	(H)(T)(T)	(T)(H)(T)

(7)	(8)
(*i*)(*ii*)(*iii*)	(*i*)(*ii*)(*iii*)
(T)(T)(H)	(T)(T)(T).

If we toss 3 coins at the same time we can get any one pair out of these 8 pairs. So the probability of getting any of these pairs is $\frac{1}{8}$. But pair (2), (3) and (4) are of the same type and so we add their probabilities getting the probability of (H H T) pair which is $\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$.

Similarly probability of (H T T) pair is $\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$.
Thus probability of H^3 is $\frac{1}{8}$.
Probability of $3H^2T$ is $\frac{3}{8}$.
Probability of $3HT^2$ is $\frac{3}{8}$.
Probability of T^3 is $\frac{1}{8}$.

Adding the left hand side we get

$$H^3 + 3H^2T + 3HT^2 + T^3 \text{ or } (H + T)^3 \qquad (3)$$

Now if we consider the three cases, (1), (2) and (3) we see that the index of $(H+T)$ represents the number of coins tossed at a time. H represents the probability of the head and T represents the probability of the tail. In formula (1) we get one chance for H and one for T. In (2) we get 1 chance for H^2, 2 for HT and 1 for T^2. This we can see by expanding $(H+T)^2$.

Similarly in (3) we get one chance for H^3, three for H^2T, three for HT^2 and one for T^3. This again we can see by expanding $(H+T)^3$. If we take more coins, say ten, we can easily see the number of chances of getting the different pairs. This we will do by expanding $(H+T)^{10}$.

After expanding $(H+T)^{10}$ by the binomial theorem we pick up the coefficients of the different combinations which represent the number of chances of their occurrence. Just as in the case of $(H+T)^3$ i.e. $(H^3+3H^2T+3HT^2+T^3)$.

Coefficient of $H^3=1$ and so we have one chance for its occurrence.

Coefficient of $H^2T=3$ and so we have three chances for its occurrence etc.

If we expand $(H+T)^{10}$ by *Binomial Theorem* we get

$$H^{10}+10H^9T+45H^8T^2+120H^7T^3+210H^6T^4+252H^5T^5+210H^4T^6+120H^3T^7+45H^2T^8+10HT^9+T^{10}.$$

Total number of happenings or pair of combinations that can occur, are $1+10+45+120+210+252+210+120+45+10+1=1024$.

Probability of getting H^{10} is $\frac{1}{1024}$ Number of favourable chances in $1024=1$

Probability of getting $10H^9T$ is $\frac{10}{1024}$ Number of favourable chances in $1024=10$

Probability of getting $45H^8T^2$ is $\frac{45}{1024}$ Number of favourable chances in $1024=45$

Probability of getting $120H^7T^3$ is $\frac{120}{1024}$ Number of favourable chances in $1024=120$

Probability of getting $210H^6T^4$ is $\frac{210}{1024}$ Number of favourable chances in $1024 = 210$

Probability of getting $252H^5T^5$ is $\frac{252}{1024}$ Number of favourable chances in $1024 = 252$

Probability of getting $210H^4T^6$ is $\frac{210}{1024}$ Number of favourable chances in $1024 = 210$

Probability of getting $120\ H^3T^7$ is $\frac{120}{1024}$ Number of favourable chances in $1024 = 120$

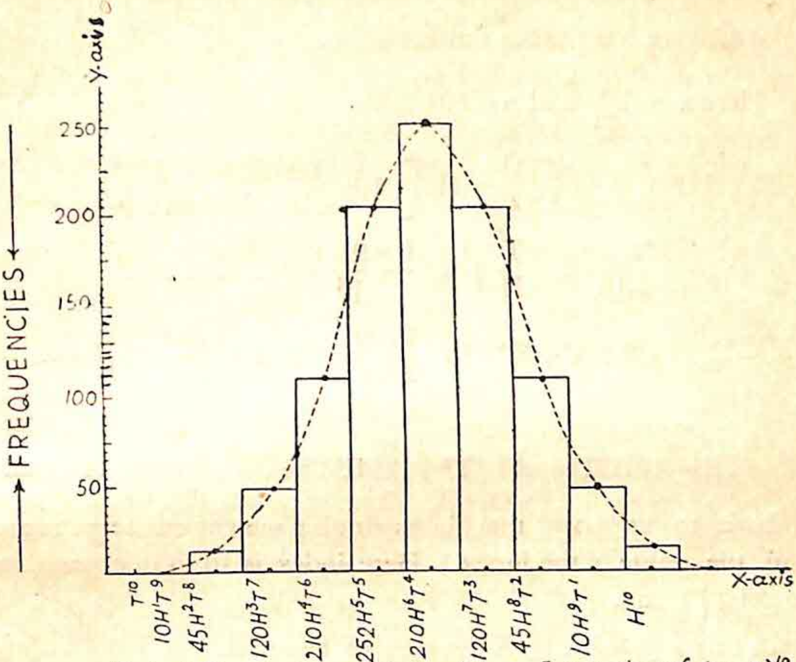
Probability of getting $45H^2T^8$ is $\frac{45}{1024}$ Number of favourable chances in $1024 = 45$

Probability of getting $10HT^9$ is $\frac{10}{1024}$ Number of favourable chances in $1024 = 10$

Probability of getting T^{10} is $\frac{1}{1024}$ Number of favourable chances in $1024 = 1$

Now we represent these graphically, taking these eleven terms of expansion on the X-axis at equal distances and chances of occurrence of the frequency on the Y-axis. We get a symmetrical frequency polygon. Again if we draw the Histogran with this data we get a symmetrical histogram.

From Data of Expansion of $(H + T)^{10}$



Probability surface obtained from Expansion of $(H+T)^{10}$

Graph No. 6.

The Binomial Theorem

Now let us discuss very briefly the Binomial Theorem.

Suppose we have to expand $(1+x)^n$ by the Binomial Theorem. The formula for the expansion on this expression is given below :

$$(1+x)^n = 1 + nx + \frac{n(n-1)x^2}{\lfloor 2} + \frac{n(n-1)(n-2)x^3}{\lfloor 3} + \ldots\ldots\ldots\ldots$$

The readers can memorize this formula and can expand any expression raised to a certain index.

Thus if we are to expand $(H+T)^{10}$ we put it in the form of $(1+x)^n$

$(H+T)^{10} = (H)^{10}\left(1+\frac{T}{H}\right)^{10}$. Now we expand

$\left(1+\frac{T}{H}\right)^{10}$ and multiply each term of the expansion by H^{10} which we have taken out common.

Here $x = \frac{T}{H}$ and $n=10$

$$\therefore\ H^{10}\left(1+\frac{T}{H}\right)^{10}=H^{10}\left[1+\frac{T}{H}\times 10+\left(\frac{T}{H}\right)^2\times\frac{10\times 9}{\lfloor 2}+\left(\frac{T}{H}\right)^3\times\frac{10\times 9\times 8}{\lfloor 3}+\ldots\ldots\ldots\ldots\right]$$

$$=H^{10}+10H^9\,T+H^8T^2\times\frac{10\times 9}{2\times 1}+H^7T^3\times\frac{10\times 9\times 8}{3\times 2\times 1}+\ldots\ldots$$

up to 11 terms.

$$=H^{10}+10H^6T^1+45H^7T^2+120H^5T^3+\ldots\ldots\ldots\text{up to 11 terms}$$

It is to be noted that in expanding we get one term more than the value of the index. Here index is 10 so in expanding we get 11 terms.

In this expansion $\lfloor\ $ is the sign of factorial.
It can be explained as below :

$\lfloor 8=8\times 7\times 6\times 5\times 4\times 3\times 2\times 1.$
$\lfloor 3=3\times 2\times 1$
$\lfloor 5=5\times 4\times 3\times 2\times 1$
$\lfloor n=(n-1)\,(n-2)\ldots 3\times 2\times 1$

or we can say $\lfloor 8=8\times 7\times\lfloor 6$ etc.

Similarly $\lfloor n=n\ \underline{|n-1}=n(n-1)\ \underline{|n-2}$ etc.

Normal Curve

The graph (Frequency polygon or histogram) which we see in the case of the data from the expansion of $(H+T)^{10}$ is a symmetrical curve. The highest measures or the frequencies fall at the centre. From this central place the measures begin to decline both in the right and left. Thus if we study this curve we see maximum measures i.e. 252 are at the centre. Then at a distance on the left and at the same distance on the right measures are 210 i.e. they have declined by the same amount and so on. If we divide this curve by a line drawn perpendicular from the highest point on the curve to X-axis we see this perpendicular divides the curve in the two parts which are equal in area. Thus we can say both the parts are symmetrical. This sort of curve is called the 'normal probability curve,' or 'normal curve.' In the case of the normal probability curve, the mean, the median and the mode, i.e., central tendencies lie at the middle point of the distribution. Thus in the above example the mean, the median and the mode will lie at a point on the X-axis which is the intersection of the X-axis and the perpendicular from the highest point of the curve on the X-axis.

Equation of the Normal Curve

The equation of the normal curve of frequency is,

$$y=y_0e^{\frac{-x^2}{2\sigma^2}}$$

where y=Ordinate representing the expected frequencies

y_0=Maximum ordinate or the ordinate at the mean. Here frequencies will be maximum.

x=Deviation from the mean in terms of standard deviation.

σ=Standard deviation of the distribution in terms of the class-interval.

e=Naperian base constant$=2{\cdot}7183$

$\pi=3{\cdot}1416$

But $y_0=\frac{N}{\sigma\sqrt{2\pi}}$ where N=Total number of cases

So the equation of normal probability curve can be written as

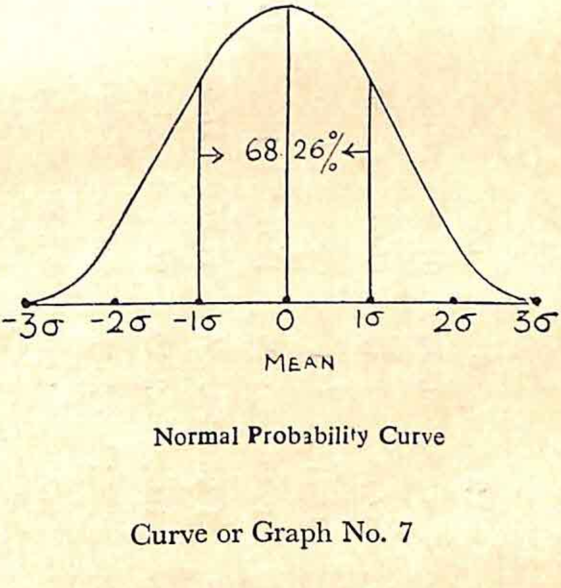
$$y=\frac{N}{\sigma\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}}$$

Tables of Areas under the Normal Curve

The total area under the normal curve is supposed to be 10000 which facilitates the calculation of the fractional parts of the area between the mean and any ordinate erected on the X-axis. Table (9) helps us to calculate the area between the mean of the normal curve and any ordinate drawn at a certain distance from the mean.

In the case of the normal curve we take deviations from the mean $(X-M)$ or x on the X-axis. If we divide this x by σ (standard deviation) then the deviation from the mean is expressed in terms of σ. Thus dividing x by σ we get $\frac{x}{\sigma}$ which means deviation x in terms of σ. Sometimes we call these $\sigma-$deviation scores as standard scores or z scores.

Now let us study table (9). Here $\frac{x}{\sigma}$ denotes the distance in tenths of σ on the X-axis from the mean of the normal curve. Distance from the mean in hundreths of σ are given by the rest of the columns. The total area of the curve represents the total number of cases. The mean of the normal curve lies at the centre of the X-axis of the normal curve. We divide the X-axis in terms of σ units say 1σ, 2σ, $3\sigma, -1\sigma$, -2σ, -3σ etc. We can find out the number of cases between the mean and any ordinate erected at various distances from the mean with the help of this table. Suppose we want the number of cases between the mean and the ordinate at 1σ. We go down the $\frac{x}{\sigma}$ column till we reach 1. Then from this point we start to the right in a row till we reach the figure 3413 under column ·00. This shows that out of 10,000 cases which are under the whole normal curve, 3413 cases lie between the mean and the ordinate at 1 σ. This means 34·13% cases lie between the mean and 1σ. Again if we want the number of cases between the mean and the ordinate at -1σ we follow the same procedure for reading the table irrespective of the signs. Thus 34·13% cases lie between the mean and ordinate at -1σ. As shown below 1 σ lies in the right of the mean



Normal Probability Curve

Curve or Graph No. 7

of the normal curve and -1 σ on the left of the mean of the curve. This means that between 1 σ and -1σ lies (34·13 + 34·13)% or 68·26% cases. Since the number of cases between the mean and 1 σ and the number of cases between the mean and -1 σ are same, we conclude that normal curve is a symmetrical curve.

(Table 9)

Fractional parts of the Total Area (taken as 10000) under the Normal probability curve, corresponding to distances on the base line between the mean and successive points laid off from the mean in units of standard deviation.

Example. Between the mean and a point $1{\cdot}38 = \left(\frac{x}{\sigma} = 1{\cdot}38\right)$ are found 41·62% of the entire are under the curve.

$\frac{x}{\sigma}$	·00	·01	·02	·03	·04	·05	·06	·07	·08	·09
0·0	0000	0740	0080	0120	0160	0199	0239	0279	0319	0359
0·1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0·2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0·3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0·4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0·5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0·6	2257	2291	2324	2357	2389	2422	2454	2468	2517	2549
0·7	2580	2621	2642	2673	2704	2734	2764	2794	2823	2852
0·8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0·9	3159	3189	3212	3238	3264	3290	3315	3340	3365	3386
1·0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621

$\frac{x}{\sigma}$	·00	0	·02	·03	·04	·05	·06	·07	·08	·09
1·1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1·2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1 3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1·4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1·5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1·6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1·7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1·8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1·9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2·0	4772	4778	4783	4788	4793	4798	4803	4805	4812	4817
2·1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2·2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2·3	4893	4896	4898	4901	4904	4906	4909	4911	3913	4916
2·4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936

$\frac{x}{\sigma}$	·00	·01	·02	·03	·04	·05	·06	·07	·08	·09
2·5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2·6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2·7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2·8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2·9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3·0	4986·5	4986·9	4987·4	4987·8	4988·2	4988·6	4988·9	4989·3	4989·7	4990·0
3·1	4990·3	4990·6	4991·0	4991·3	4991·6	4991·8	4992·1	4992·4	4992·6	4992·9
3·2	4993·129									
3·3	4995·166									
3·4	4996·631									
3·5	4997·674									
3·6	4998·409									
3·7	4998·922									
3·8	4999·277									
3·9	4999·519									
4·0	4999·683									
4·5	4999·966									
5·0	4999·997133									

[Table 10]

Fractional parts of the total area (taken as 10000) under the normal probability curve, corresponding to distances on the base line between the mean and successive points laid off from the mean in units of PE.

Example. Between the mean and a point 1·55 PE$\left(\frac{x}{\text{PE}}=1{\cdot}55\right)$ from the mean are found 35·21% of the entire area under the curve.

$\frac{x}{\text{PE}}$	·00	·05	$\frac{x}{\text{PE}}$	·00	·05
·0	0000	0135	3·0	4785	4802
·1	0269	0403	3·1	4817	4832
·2	0537	0607	3·2	4846	4858
·3	0802	0933	3·3	4870	4881
·4	1063	1193	3·4	4891	4900
·5	1320	1447	3·5	4909	4917
·6	1571	1695	3·6	4924	4931
·7	1816	1935	3·7	4937	4 43
·8	2053	2168	3·8	4984	4953
·9	2281	2392	3·9	4957	4961
1·0	2500	2606	4·0	4965	4968
1·1	2709	2810	4·1	4972	4974
1·2	2909	3004	4·2	4977	4979
1·3	3097	3187	4·3	4981	4983
1·4	3275	3360	4·4	4985	4987
1·5	3442	3521	4·5	4988	4989
1·6	3597	3671	4·6	4990	4961
1·7	3742	3811	4·7	4992	4993
1·8	3876	3939	4·8	4994	4995
1·9	4000	4058	4·9	4995	4996
2·0	4114	4166	5·0	4996	499
2·1	4217	4265	5·1	4997·1	4994
2·2	4311	4354	5·2	4997·7	4998
2·3	4396	4435	5·3	4998·2	4998·5
2·4	4473	4508	5·4	4998·6	4998·8
2·5	4541	4573	5·5	4999	4999·1
2·6	4603	4631	5·6	4999·1	4999·3
2·7	4657	4682	5·7	4999·4	4999·5
2·8	4705	4727	5·8	4999·54	4999·6
2·9	4748	4767	5·9	4999·65	4999·7

Again, let us find out the number of cases between the mean and 2 56 σ. For this we go down $\frac{x}{\sigma}$ column till we reach 2·5 and then move to the right in a row till we reach 4948 below the column ·06. This means that 4948 cases out of the entire 10000 cases lie between the mean and the ordinate at 2·56 σ. or 49·48% cases lie between the mean and the ordinate at 2·56 σ. By symmetry 49·48% cases also lie between the mean and $-2{\cdot}56\ \sigma$. Thus 98·96% cases lie between $\pm 2{\cdot}56\ \sigma$. Now let us find out the number of cases between 3 σ and $-3\ \sigma$ or $\pm 3\ \sigma$. For this we consult the table 9 and see 4986·5 cases lie between the mean and 3 σ, i.e. 49·865% cases lie between the mean and 3σ. By symmetry 49·865% cases also lie between the mean and -3σ. Thus 99·730% cases lie between $\pm 3\ \sigma$. This means almost cent-percent cases lie between $\pm\ 3\sigma$ in a normal curve. Thus we suppose that the normal curve terminates on the X-axis at 3 σ on the right and at $-3\ \sigma$ on the left though actually it never meets. Only negligible number cases lie beyond these limits in a normal curve.

Thus we see, table 9 helps us in calculating the cases between the mean and various ordinates erected at various points on the X-axis from the mean.

Area in terms of PE units

As we have read in the chapter (measures of variability) the quartile deviation

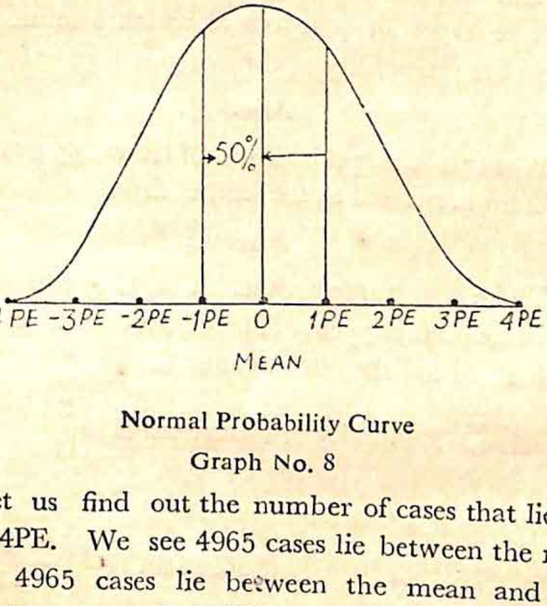
$$Q = \frac{Q_3 - Q_1}{2}$$

Where Q_3 = Third quartile
Q_1 = First quartile.

In the case of normal curve, quartile deviation Q is called as PE. When we want to find out the area or the number of cases under the normal probability curve, corresponding to distances on the base line between the mean and successive points laid of from the mean in units of PE we consult the table (10). Here x is the deviation from the mean. We divide it by PE.

getting the deviations in terms of PE. This table cannot read so accurately as table 9. It can read the distances only in tenths and 0·5th. If we want smaller divisions we can use interpolation.

The procedure for consulting this table is the same as in the case of table 9. If we want the number of cases between the mean and + 1 PE we go down the column $\frac{x}{\text{PE}}$ till we reach and then proceed to the right in a row till we reach the figure 2500 under column ·00. This means 2500 cases out of the entire 10000 cases lie between the mean and the ordinate at 1 PE. By symmetry 2500 cases lie between the mean and the ordinate at -1 PE. Thus 5000 cases i.e. 50% cases lie between $+1$ PE and -1 PE or between ± 1 PE.



Normal Probability Curve

Graph No. 8

Again let us find out the number of cases that lie between 4 PE and -4PE. We see 4965 cases lie between the mean and 4 PE. Also 4965 cases lie between the mean and -4 PE. Thus 9930 cases out of 10000 or 99·3% cases lie between $\pm$4PE. Thus we can say that approximately cent-percent cases lies between $\pm$ 4 PE and we assume the normal curve terminates on the X-axis at a distance 4 PE on the right and -4 PE on the left. Of course a few cases lie beyond these limits too.

But for our purpose we will consider the terminating limits of normal curve as ± 4 PE from the mean on the X-axis.

EXERCISES

1. What is the chance of throwing a number greater than 4 with an ordinary die whose faces are numbered from 1 to 6 ?

Solution. There are 6 possible ways in which the die can fall and of these two are favourable to the event required. Therefore the required chances

$$=\tfrac{2}{6}=\tfrac{1}{3} \qquad \textbf{Ans. } \tfrac{1}{3}.$$

2. Two coins are thrown at the same time. What is the probability of throwing at least one head ?

Solution. When we throw two coins we can get both heads, or both tails or one head one tail or one tail one head. Thus there are 4 types of pair out of which 3 contains at least one head. So the probability of throwing at least one head $=\tfrac{3}{4}$.

Ans. $\tfrac{3}{4}$.

3. What is the probability of throwing exactly one head if three coins are tossed at the same time.

Ans. $\tfrac{3}{8}$.

4. There are three events A, B, C ; one of which must and only one can happen. The odds are 8 to 3 against A, 5 to 2 against B. Find the odds against C.

Solution. Possibility of failure for $A=\frac{3}{11}$

" " " $B=\tfrac{5}{7}$

" " Success for $A=1-\frac{8}{11}=\frac{8}{11}$

" " " $B=1-\tfrac{5}{7}=\tfrac{2}{7}$

" " Success for C

$$=1-\left(\frac{3}{11}+\frac{2}{7}\right)=\frac{34}{77}=\frac{34}{34+43}$$

Possibility of failure for C$=\frac{4}{34+43}$

or odds are 43 to 34 **Ans.**

5. Find the chance that if a card is drawn at random from an ordinary pack, it is one of the court cards.

Solution. Since there are 52 cards in a pack, the number of ways in which a card can be drawn=52.

The number of favourable ways=12 (Since king, queen and knave of each colour is to be included in the court-cards)

$$\therefore \text{ Chance}=\frac{12}{52}=\frac{3}{13} \textbf{ Ans.}$$

6. Five coins are thrown. What is the probability that exactly two of them will be heads.

$$\textbf{Ans. } \frac{10}{32}$$

7. What percentage of the normal distribution lie between

(a) $\pm 2{\cdot}22\ \sigma$

(b) $\pm 2{\cdot}5$ PE

(c) mean and $-2{\cdot}6\ \sigma$

(d) mean and $1{\cdot}8$ PE

(e) $-1{\cdot}2\ \sigma$ and $2{\cdot}5\ \sigma$.

(f) $-1{\cdot}3\ \sigma$ and $-2{\cdot}2\ \sigma$.

Ans. (a) 97·36 %

(b) 90·82 %

(c) 49·53 %

(d) 38·76 %

(e) 87·87 %

(f) 8·29 %.

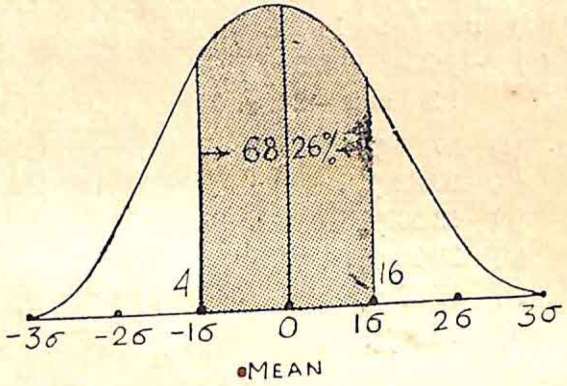
CHAPTER 21

APPLICATION OF THE NORMAL PROBABILITY CURVE

WITH THE HELP OF TABLES 9 AND 10 WE CAN SOLVE A number of problems related to the normal probability curve. By utilizing the mathematical properties of normal distribution it is possible among other things to determine difficulty of test items, to determine the extent to which a particular group of scores represents a normal distribution.

Below are given a number of examples which the students can carefully work out themselves.

Example 1. The mean of a normal distribution is 10 and the SD is 6. Calculate the percentage of the cases that will lie between 4 and 16.

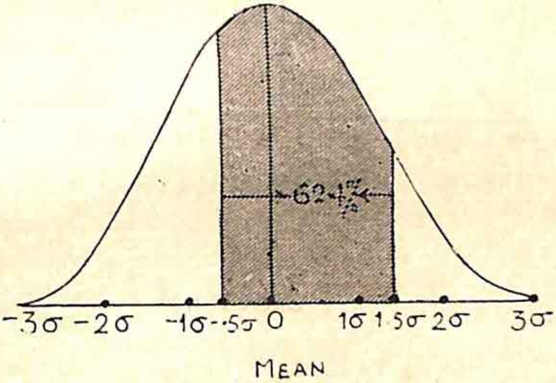


Curve No. 9

Here the mean is 10. We see 16 is 6 points above the mean and similarly 4 is 6 points below the mean. Thus in the former case deviation from mean i. e. $x=6$ and also in the latter case $x=6$. To plot these x^s on the base line in terms of σ we divide them by σ separately.

Thus 16 is 1 σ above the mean and 4 is 1 σ below the mean as shown in the [curve 9]. Now our problem reduces to finding out the number of cases between $\pm$ 1 σ. We consult [table 9] and see that between $\pm$ 1 σ, 68·26 % cases lie. Thus we can say between 4 and 16 lies 68·26 % cases. Shaded area contains 68·26 % cases.

Example 2. The mean of a normal distribution is 10 and SD is 4. What percentage of the cases lie between 8 and 16 ?



Curve No. 10

In this case, mean$=10$ and $\sigma=4$.

8 is 2 points below the mean or $\frac{2}{4}$ σ or $\frac{1}{2}$ σ or ·5 σ below the mean.

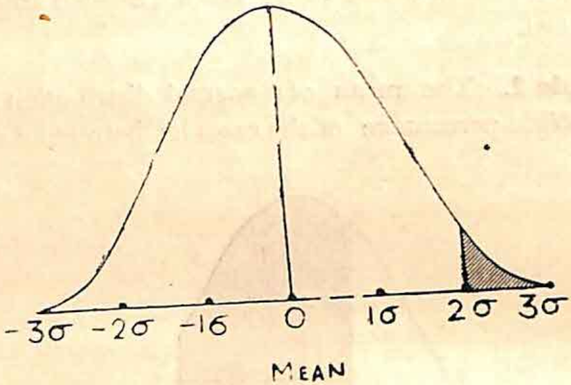
16 is 6 points above the mean or $\frac{6}{4}$ σ or 1·5 σ or 1·5 σ above the mean. That means we are to find out the number of cases between $-$·5 σ and 1·5 σ.

Consulting table 9 the number of cases between the mean and $-$·5 σ are 19·15 % Again number of cases between the mean and 1·5 σ are 43·32 %. Thus the total number of cases between $-$·5 σ and 1·5 σ or between 8 and 16 is

$$[19{\cdot}15+43{\cdot}32]\ \% = 62.47\%$$

Shaded area represents 62·47 % cases.

Example 3. In a normal distribution mean $=12$ and $\sigma=8$. What percentage of the scores lie above 28 ?



Curve No. 11

28 is 16 points above the mean i.e. $\frac{16}{8}$ σ or 2 σ above the mean. We know that mean divides the normal distribution in two equal parts each of which contains 50% scores. Thus half of the normal distribution on the right has 50% scores. We are to find the scores that lie beyond 2 σ on the right. For this we find the percentage of the scores that lie between the mean and 2 σ and subtract this percentage from 50%.

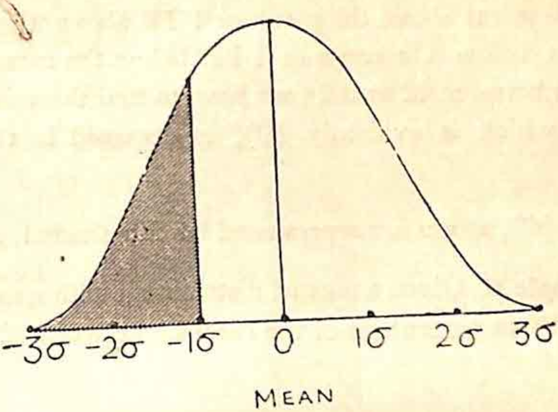
From table 9 percentage of scores between the mean and 2 $\sigma=47{\cdot}72$.

$\therefore$ $[50-47{\cdot}72]\ \%=2{\cdot}28\%$

This means 2·28% cases lie above 2 σ or 16.

Shaded area contains these 2·28 % scores.

Example 4. In a normal distribution mean $=12$ and $\sigma=8$. What percentage of the scores lie below 4 ?



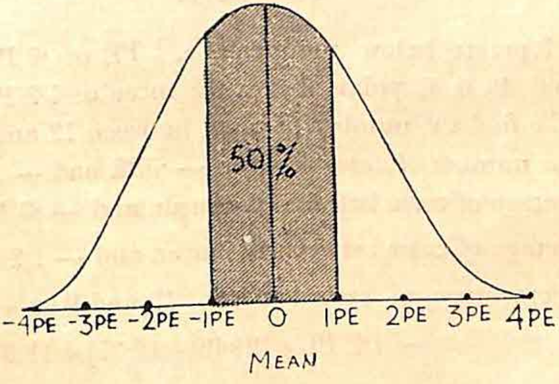
Curve No. 12

Here 4 is 8 points below 12 [the mean] i.e. 1 σ below the mean. On the left half of the normal distribution we have 50% scores. Between the mean and $-1\ \sigma$ we have 34·13% scores.

$\therefore$ Below $-1\ \sigma$ or 4 points we have $(50-34{\cdot}13)\%$ or 15·87% scores.

The reasoning is same as in example 3. These 15·87% scores lie in a shaded area.

Example 5. In a normal distribution mean is 18 and $Q=PE=6$. What percentage of the scores lie between 12 and 24 ?



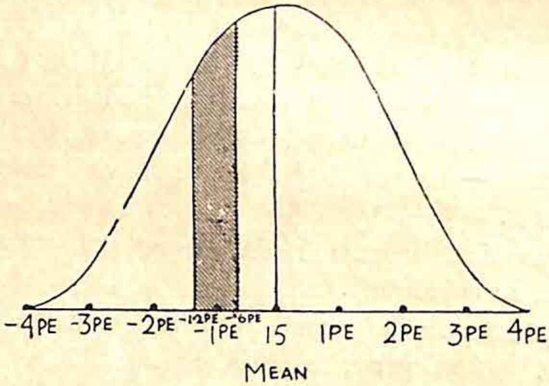
Curve No. 13

In this case the mean = 18 and PE = 6.

24 is 6 points above the mean or 1 PE above the mean. 12 is 6 points below the mean or 1 PE below the mean. To find the scores between 12 and 24 we have to find the scores between ±1 PE, which is evidently 50% as discussed in the previous chapter.

These 50% scores are represented by the shaded area.

Example 6. Given a normal distribution with a mean 15 and PE = 5. What percentage of the cases lie between 12 and 9 ?



Curve No. 14

12 is 3 points below the mean i.e. $\frac{3}{5}$ PE or ·6 PE below the mean and 9 is 6 points below the mean or 1·2 PE below the mean. To find the number of cases, between 12 and 9 we have to find the number of cases between — ·6PE and — 1·2 PE.

Percentage of cases between the mean and — ·6 PE = 15·71.

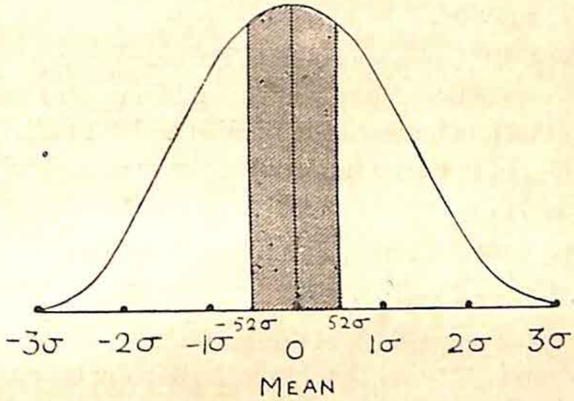
Percentage of cases between the mean and — 1·2 PE = 29·09.

∴ Percentage of cases between 12 and 9 or — ·6 PE and

$$-\ 1{\cdot}2\ \text{PE} = [29{\cdot}09 - 15{\cdot}71] = 13{\cdot}38$$

Shaded area represents these 13·38 % cases.

Example 7. In a Normal Distribution mean is 14 and $\sigma=6$. What are the limits of the distribution which contains 4% middle cases ?



Curve No. 15

The mean$=14$ and $\sigma=6$.

Out of these 40% cases, 20% will lie on the right half of the distribution and rest 20% on the left half.

20% cases means 2000 cases out of 10000. From table 9 we see 1985 cases lie between the mean and $\cdot 52\ \sigma$. This means approximately 20% cases lie between the mean and $\cdot 52\sigma$. Similarly on the left hand side 20% cases lie between the mean and $-\cdot 52\ \sigma$. Thus 40% middle cases lie between the limits $\cdot 52\ \sigma$ and $-\cdot 52\ \sigma$. But $\sigma=6$ (given).

$\therefore$ 40% middle cases lie between $\cdot 52\times 6$ and $-\cdot 52\times 6$ or $3\cdot 12$ and $-\ 3\cdot 12$ from the mean. The value of mean is 14.

Therefore 40% middle cases lie between 17·12 and 10·82.

Example 8. In a normal distribution median is 150 and PE$=20$. What limits will include the highest 40% of the scores ? What limits will include the lowest 20% of the scores ?

The highest 40% scores of a normally distributed group will have 10% of the cases between its lower limit and median. From Table 10 we see 10·63% cases lie between the median and ·4 PE or approximately 10% cases lie between the median and ·4 PE. But PE is 20. ∴ ·4 PE=·4×20=8 points above the median. But median=150.

Therefore 8 points above median means 150+8 or 158.

The lower limit of the highest 40% scores of the given group is 158 and the upper limit is the highest score in the distribution whatever that may be.

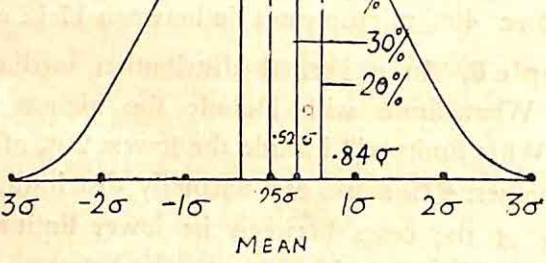
The lowest 20% of a normally distributed group will have 30% of the cases between the median and its upper limit. 3004 cases out of 10000 lie between the median and 1·25 PE.

That is 30% of the cases lie between the median and — 1·25 PE. But PE=20.

∴ — 1·25 PE= — 1·25×20 points or 25 points below the median i. e. 150 — 25=125.

Therefore, the upper limit of the lowest 20% of the scores in the given group is 125 and the lower limit is the lowest score in the distribution.

Example 9. In a group of boys 20% could solve a problem. Another problem was solved by 30% boys of the same group and the third problem was solved by 40% boys of the same group. Assuming capacity measured by test problems to be distributed normally, what is the relative difficulty of questions (1), (2) and (3) ?



Curve No. 16

The first step is to find the position in the distribution such that 20% cases lie above it and 80% cases lie below it because in solving the first question 20% boys passed and 80% failed.

Again we know the normal distribution is divided by the mean in two equal parts.

In each part lies 50% cases. To find the point above which lie 20% cases means, to find the point between which and the mean lie 30% cases. Using table (9) 29·65% cases lie between the mean and ·84σ. It means approximately 30% cases lie between the mean and ·84 σ. Thus question one belongs to a point on the base line of a curve, a distance ·84 σ from the mean and accordingly ·84 σ may be set down as difficulty value of this question.

In the second question 30% boys passed and 70% failed. To find the point in the distribution above which lie 30% cases means, to find the point between which and the mean lie 20% cases.

From Table 9 approximately 20% cases lie between the mean and ·52 σ. Hence question second has a difficulty value at the point ·52 σ.

Again for the third question we find the point between which and the mean lie 10% of the cases. Evidently between the mean and ·25 σ lies 10% cases and thus at ·25 σ is the difficulty value of the third question.

So the first, second and the third questions have difficulty values at ·84 σ, ·52 σ and ·25 σ respectively.

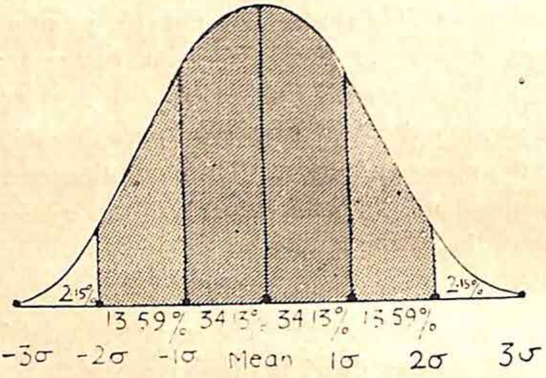
The σ difference between (2) and (3) is ·27

The σ difference between (1) and (2) is ·52

Thus first σ difference is approximately $\frac{2}{3}$ of the second σ difference. Since the percentage difference is the same in two comparisons, it is evident that when ability is assumed to follow the normal distribution, σ and not percentage differences are the better indices of differences in difficulty.

The Percent of the Distribution included in each of the First Three σs.

The normal curve below exhibits the distribution included in each of the first three σs. It is advisable that the readers may memorize them and consult the table to varify the truth.



Curve No. 17

Thus total number of cases inside the whole normal distribution,

$$=(2{\cdot}15+13{\cdot}59+34{\cdot}13+34{\cdot}13+13{\cdot}59+2{\cdot}15)\%$$
$$=99{\cdot}74\%.$$

Important Characteristics of a Normal Distribution

(1) The first SDs in either direction from the mean of a normal distribution each include 34·13 percent of the distribution of cases (as might be estimated from the areas) included under the curve when perpendiculars are erected at 1 SD points above and below the mean or at the + 1 SD and − 1 SD points.

(2) Succeeding SDs however, include much smaller percents of the distribution of cases, the second SDs including 13·59 percent each and the third SDs only 2·15 percent. SDs are equal, therefore, in score units or points but these include different percents of total number of cases or measurements of a distribution when perpendiculars are erected at the points which separate them.

(1) The percent of the scores included between the mean and a perpendicular erected at any SD or fractional SD point from the mean is always constant i.e. perpendiculars erected at ·6 of SD or 1·7 SD from the mean will always include the same percent of the normal distribution between these points and the mean.

(4) From the above point (3) just presented, it follows that given the percent of cases between the mean and a perpendicular erected at a point some distance from the mean we should be able to find the distance in terms of SD units, and given the distance between the mean and a point located some distance from it in terms of SD units, we should be able to tell the percent of the distribution which lies between the mean and a perpendicular erected at a given point. In order to translate percent of this distribution into SD units and vice versa, it is necessary to work with an SD table, which gives the equivalent of each in terms of the other.

Example 10. There are 100 students in a class. You are to grade them in 6 groups A, B, C, D, E, F, according to the ability.

The range of ability of groups A, B, C, D, E, F is same.

Assuming that the behaviour measured by your test is normally distributed, how many students should be placed in groups A, B, C, D, E and F ?

Let us follow the curve No. 17. In this case we have 6 groups of students and we are to represent their positions on the curve. We know that our base line extends from $-3\,\sigma$ to $3\,\sigma$ i.e. its length is $6\,\sigma$ and over this distance we are to accomodate 6 groups each having the same range of ability. We divide $6\,\sigma$ by 6 (number of groups) which gives σ. This is the range of each group. Thus our base line has been divided into 6 parts. The perpendiculars erected on the base lines at $-3\,\sigma$ $-2\,\sigma$, $-\,\sigma$, mean, σ, $2\,\sigma$, $3\,\sigma$ divides our distribution in 6 groups. It is easy for us to calculate the number of cases in these groups. As shown in curve (17) 2·1 % cases lie between 2 σ and 3 σ because 47·72 percent cases lie between the mean and 2 σ. Again 49·866 percent cases lie between the mean and 3 σ. Thus (49·87—47·72)% or 2·15% cases lie between 2 σ and 3 σ. Similarly 13·59% students lie between +1 σ and +2 σ. 34·13% students lie between the mean and +1 σ. 34·13% students lie between the mean and −1 σ. 13·59% students lie between − σ and − 2 σ, and 2·15% students lie between −2 σ and −3 σ.

In short.

GROUPS

	A	B	C	D	E	F
% of the total in each.	2·15	13·59	34·13	34·13	13·59	2·15
Number in each	2	14	34	34	14	2

Total = 100 students

This a very important method for a class teacher to grade his students according to their abilities.

Example 11. A class-teacher classifies 100 students into groups A, B, C, D, E. The range of the ability of each group is same. If the trait measured by the test is normally distributed how many students may fall in each group ?

Hints. Students may try this question themselves with the help of the following hints.

Here range of the base line is 6 σ and we have divide it into 5 parts of the same range. So the range of the individual group is 1·2 σ. Thus group A falls between 1·8 σ and 3 σ. Between 1·8 σ and 3 σ percentage of students = 49·86 − 46·41 = 3·5

Group B falls between ·6 σ and 1·8 σ

where percentage of students = 23·84

Group C falls between ·6 σ and −·6 σ

where percentage of students $=(22{\cdot}57+22{\cdot}57)$
$=45{\cdot}14.$

[Because half of the portion is on the right of the mean and half on the left.]

Group D falls between $-{\cdot}6\,\sigma$ and $-1{\cdot}8\,\sigma$
and percentage of students in this part $=23{\cdot}84$

Group E falls between $-1{\cdot}8\,\sigma$ and $-3\,\sigma$
and percentage of students in this part $=3{\cdot}5$

Thus number of students in A=4 or 3 ; in B=24 ; in C=45 in D=24 and in E=4 or 3.

EXERCISES

1. Given a normal distribution with a mean of 14 and $\sigma=6$. What percentage of the cases falls between 8 and 20 ?
Ans. 68·2 %

2. Given a normal distribution with a mean of 12 and $\sigma=4$. What percentage of the cases lies above 18 ?
Ans. 6·68 %

3. Given a normal distribution with mean=18 and PE=8. What percentage of the scores lies between 10 and 38 ?
Ans. 70·41 %

4. In a sample of 1000 cases the mean of a certain test is 14·40 and $\sigma=2{\cdot}50$. Assuming normality of the distribution, how many individuals score between 12 and 16 ?

Hints :—12 is ·96 σ below the mean and from the table 33·15 % cases lie between the mean and ·96 σ.

16 is ·64 σ above the mean and from the table 23·89 % cases lie between the mean and ·64 σ. Thus in all 57·04 % cases lie between 12 and 16.

Infinite Population

An infinite population is one which contains as infinite number of members. For example the population of the atmospheric pressure at various points, temperature at various places etc.

Population of Populations

A population contains a number of members just as the population of India but again this population is a member of a more wider population, e. g., population of the world. Thus this population of the world is the *population of populations.*

Types of Sampling

There are three types of sampling :

(1) *Random sampling,*
(2) *Purposive sampling,*
(2) *Mixed sampling.*

An illustration will distinguish between these three types of sampling. Suppose we are to take a sample of boys from a school to get an idea of their health. For this we can select the boys in the following three ways :

(1) We can select the boys at random from the list, i.e., the health report of boys. This is called a *Random sampling.*

(2) We may divide the boys into different age-groups and calculate the average health of each group. A boy in each group whose health resembles with the average health of that group may be taken. This is the *Purposive sampling.*

(3) We may divide the boys into a few age-groups and out of these groups we may at random select 10 boys from each. This is called the *Mixed sampling* because it is a combination of the first and the second types.

Human Bias

It is fact that human being is a very poor instrument for the conduct of random selection. Wherever there is any choice on the part of an observer bias is to creep in random sampling. In the experiment carried out some years back a set of eight wheat plants were selected. Six of these shoots were chosen by purely Random-Methods. The other two were chosen 'at random' by the eye. If in any set the eight shoots were ranged in the order of the magnitude, the two chosen by the eye could have any place from one to eight, and if they in common with the other six, were really random, they should have occupied these places with equal frequency in a reasonably large number of sets.

Conditions of Sampling

For a good sample the following conditions are necessary :

(1) **Independence**—The various members of the sample must be independent. Each member should have the same chance for being selected. The selection of one member should not affect the chances of the other member for being selected.

(2) **Homogeneity**—There should not be any difference in the locality from which observations have been taken. At all if there are differences in these localities they should not be of such type which may affect the behaviour of our observations. Thus the observations of the rain fall collected from one locality will not be homogeneous with those collected from any other locality. The homogeneity in behaviour is also lost due to the lapse of time. If we collect the observations for rainfall at one place to-day and do the same after a lapse of time certainly our behaviour of the observations will be changed.

Random sample and Tippetts Random Numbers

As discussed before whenever selection is made haphazardly, i. e., at random we talk of random sample There are so many difficulties in having suitable methods for getting random sample. To avoid these difficulties certain tables of random sampling numbers have been constructed. Tippett constructed one such table which consists of 41600 digits taken from a census report and combined by fours to give 10,400 four figured numbers. Some such forty numbers are given below for illustration.

2952	6641	3992	9792	7979	5911	3170	5624
4167	9524	1545	1396	7203	5356	1300	2693
2370	7483	3408	2762	3563	1089	69[illegible]3	7691
0560	5246	1112	6107	6008	8126	4233	8776
2754	9143	1405	9025	7002	6111	8816	6446

Difference between Simple Sampling and Random Sampling

Simple sampling is a Random sampling but in the former case each member of the population has the same chance for success. Say when we toss a coin the chance of getting head is ½ and chance of getting tail is ½. If we again toss it, the chance of getting head is ½ and chance of getting tail is ½. If we go on tossing the coin (assuming the coin will not be destroyed) each time head has the same chances as the tail. The occurrence of head in the first trial will not affect the chances for tail in the next trial or the occurrence of tail in the first trial will not affect the chances for the head in the next trial. In each trial head and tail will have equal chances of occurrence. Here coin is tossed at random and the chances of head and tail are equal in each case. So it is simple sampling.

The difference between Simple and Random sampling will be clear from the following example. Suppose out of 20 women and 10 men we select, at random, a sample of two. At the first moment the chance for a woman to be selected is $\frac{20}{30}$ and the chance for a man to be selected is $\frac{10}{30}$.

Suppose this time we could select one man and one woman. In our next trial there are 19 women and 9 men. So this time the chance for the selection of a woman is $\frac{19}{28}$ and the chance for the selection of a man is $\frac{9}{28}$. Thus the latter chances are affected by previous selection. The sampling is random but not simple. Thus we can differentiate between Simple and Random sampling.

Stratified Sample

If the population is divided into 'strata' by purposive methods and then a portion of sample is taken from each 'stratum', the sampling is said to be '*Stratified*'. An example of this type has been given under head 'mixed sampling'. In practical problems we generally use this method.

Controlled Sample

A controlled sample is one in which the selection is not left to chance or not entirely to chance, but in which the distribution of some selected characteristic is made to conform to some predetermined proportion. *Controlled Sampling* may or may not involve random sampling.

Matched or Equated Sample

Matched or equated samples represent the special case of controlled sampling in which two or more similarly controlled samples are drawn from the same population.

CHAPTER 23

RELIABILITY AND TEST OF SIGNIFICANCE

The meaning of Reliability

THE TRUE MEAN OR THE TRUE STANDARD DEVIATION CAN BE obtained by taking into account the scores made by all the members of the population. But to save the time and labour we draw a sample from the population and obtain the mean and the standard deviation of the sample presuming that these sample measures would be the representative measures of the population. We expect the behaviour of this sample to be the same as the behaviour of the population from which it is drawn. But we should note that the slight differences in the composition of the sample affects the value of these measures. The mean or the standard deviation or some other measures of the sample are always the 'estimates' of their population counterparts. The more this estimate is accurate, the more is our statistics reliable. We cannot obtain the true measures but we can compute the limits within which the true measures may fall. This computation of the limits is discussed in the later part of this chapter. If the estimated measure falls within this range of true measure we say with some confidence that it is reliable.

THE RELIABILITY OF THE MEASURES OF CENTRAL TENDENCY

The Reliability of the Mean

As has been said earlier that to obtain the true measures we should deal with the whole population and not the sample of the population. But at times it is difficult to deal with the whole population and under such circumstances we have to be satisfied with our operations on the sample only. However, we should note that the sample selected should be representative of the whole population and should not be a select one. It should represent the behoviour of the population as closely as possible. One thing more should be kept in view and it is that the sample should be as large as much possible because the larger is the sample the nearer is the estimate to the true measure.

Mathemetically we can show that the reliability of the mean depends upon two factors :

(1) Number of cases (N)

(2) The variability of the measure.

We shall explain both these factors.

(1) We know that the mean is the sum of all the scores divided by the total number of scores. If we add one measure or score more it affects the mean unless it coincides with the mean. For example the mean of 2, 3 and $4 = \frac{9}{3} = 3$; but if we add one score more to this say 5 then the new mean becomes

$$\frac{2+3+4+5}{4} = 3{\cdot}5.$$

If we add the mean score 3 to 2, 3 and 4 and then find out the mean again, the mean value remains the same and does not change as shown below :

$$\frac{2+3+4+3}{4} = \frac{12}{4} = 3.$$

If the number of scores is large then the addition or subtraction of one score does not influence the stability of the mean as much as it may do when the number of scores is less. For example the mean of 2, 3, 4 is 3 and the addition of one single score 5 affects the mean. The difference, between the first mean and the second mean is quite significant. If we take the mean of large number of cases say 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, it will be $\frac{465}{30} = 15{\cdot}5$. Here if we add a score five to these scores and then find out the mean the new mean will be 15·2 (approximately). Here we see a lesser difference between the means. If the number ofscores be very large, say 1000, in that case the addition or subtraction of one score will not affect the mean much. The reliability of mean increases not in proportion to the number of measures but in proportion to the square root of the number of measures. For example the mean obtained from 81 scores will be 3 times as reliable as obtained from 9 scores because

$$\frac{\sqrt{81}}{\sqrt{9}}=\frac{9}{3}=3.$$

(2) If the standard deviation of the distribution is large, the individual score tend to scatter widely and we cannot locate those cases which we have not measured. We cannot say whether those cases which we have not measured are close to the mean or far off from it. If however, the standard deviation is small we can say with certainty that those cases which we have not measured will fall close to the mean. The reliability of the mean, thus, depends upon the S. D. As S. D. increases, the reliability decreases and the vice versa.

The standard error of the mean is denoted by the following formula.

SE_{mean} or $\sigma_M = \frac{\sigma}{\sqrt{N}}$. This formula is used when the number of cases is large, say more than 50 or so. An increase in the value of σ increases the value of σ_M (standard error of the mean). A decrease in the value of N also increases the value of σ_M. The increase in the value of σ_M decreases the reliability of the mean and vice versa. Reliability of mean increases if the value of σ_M decreases, i.e., when σ decreases and N increases.

Example (1)—The mean achievement score of 6400 eleven year boys of U. P. in an achievement test is 65·3, with a σ of 3·2. How reliable is this mean? How much does it probably diverge from the mean which would have been obtained had all the eleven year boys in the province been measured?

Solution. The sample mean is 65·3. We could have taken other samples at random. It is certain that in doing so we could have obtained some other values of the mean and 65·3 would have been one of them. With the help of the sampling theory we can calculate the probable range within which all of these possible sample means lie. It can be proved experimentally that the sampling fluctuations around the population mean follow the normal probability curve when the samples are random. The 'standard error' of the mean, i.e., σ_M, is the standard deviation of the means which would be obtained by taking all possible samples. If we are wrong five times in 100 trials, we can put the lowest obtained mean from a sample at $(65{\cdot}3 - 1{\cdot}96\,\sigma_M)$ and the highest mean obtainable at $(65{\cdot}3 + 1{\cdot}96\,\sigma_M)$. If we are wrong only once in 100 trials, we can put the lowest obtained mean at $(65{\cdot}3 + 2{\cdot}58\,\sigma_M)$ and the highest mean at $(65{\cdot}3 + 2{\cdot}58\,\sigma_M)$. For assigning these limits we consult table 9. From the table we see that 95% of the cases in a normal distribution fall between the limits $\pm 1{\cdot}96\,\sigma_M$ (5% lying outside these limits) and 99% of the cases fall between the limits $\pm 2{\cdot}58\,\sigma_M$ (1 % lying outside these limits).

Applying the formula for standard error we get

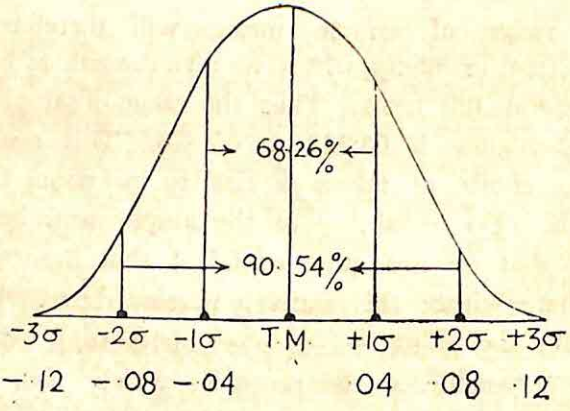
$$\sigma_M = \frac{3{\cdot}2}{\sqrt{6400}} = {\cdot}04.$$

The range of sample mean, will therefore, be 65·3 $\pm$ $(1{\cdot}96 \times {\cdot}04)$ or $65{\cdot}3 \pm {\cdot}078$ if we take the risk of being wrong five times in 100 trials. Thus the range of sample means for lowest to highest is 65·222 to 65·378. This range depends upon our choice of taking a risk to go wrong five times in 100 trials. The reliability of the sample mean depends upon the fact that we are quite confident that the true mean lies some where within this relatively narrow range. If we take a lesser risk, say going wrong once in 100 trials we can predict with more confidence the range in which the true mean probably lies. That range will be $65{\cdot}3 \pm (2{\cdot}58 \times {\cdot}04)$ or $65{\cdot}3 \pm {\cdot}103$ or 65·197 to 65·403. Since the range within which the true mean probably falls is quite narrow in either case, we conclude that our obtained mean cannot be very far-off from the true value and the considerable confidence may be placed in its adequacy.

This problem will be easily followed if solved in the following way. Let us suppose that we take 100 random samples each containing 6400 examinees. If we find out the mean of each sample it will vary to some extent from the true mean or the mean of the population. So we will get 100 means. There will be a large deviation in only a few cases, small deviation in most of the cases and zero deviation in many cases. Even if the samples are not normally distributed, their means will do so.

The means of all these estimates or sample means will be the best estimate of the true mean. Again the standard deviation of these 100 means will be the standard error of the mean which we have calculated, i. e., ·04. We can say σ_M measures the spread of these means around the population mean or true mean.

As shown in the following graph the 100 means are represented by a normal frequency distribution around TM and σ as ·04. The ordinate y represents the frequency of each mean. Since the ordinate at TM is maximum we can see



Curve 18

Sample distribution of means showing variability of obtained means around the True or population Mean (TM) in terms of σ_M (·04).

that the true mean occurred most of the time. We can see from the table that between $\pm 1\sigma$, 68·26% cases lie. Therefore, we can say 68 sample means out of 100 are expected to differ from TM by not more than $\pm 1\sigma$ (·04) and about 96 of our obtained sample means are expected to differ from TM by not more than $\pm 2\sigma$ (·08). Since our mean 65· 3is one of these sample means, we can say that the probability is approximately ·96 that 65·3 does not miss the true mean by more than ± ·08.

Determining limits of accuracy

We are now in a position to see that the reliability depends upon the being of our estimate closer to the parameter or the true measure. The question arises how closer. For this reason we fix a certain limit within which our obtained measures or estimates should lie for being called reliable. Fisher has suggested two accuracy limits called ·05 level and ·01 level respectively. From table (9) we see that 95 % cases lie within the limit $\pm 1{\cdot}96\ \sigma_M$. Hence odds are 95 : 5 or 19 : 1 that any sample mean will lie within these limits. Again 99 % cases lie within the limits $\pm 2{\cdot}58\ \sigma_M$, the odds are 99 : 1 that any sample mean will not miss the true mean by more than $\pm 2{\cdot}58\ \sigma_M$. In the above problem we can say with confidence (the odds are 19 : 1) that the sample mean 65·3 does not miss the true mean by more than ±·078. We can say with more confidence (odds are 99 : 1) that the sample mean 65·3 does not miss the true mean by more than ±·103. The ·01 level is more significant than ·05 level.

The SE of the Mean in Small Samples

There is a difference between the σ of the sample and the σ of the population. The σ of the sample is always less than the σ of the population. The smaller is the sample the lesser is its σ than the σ of the parent population. To correct the sample σ we should calculate it with the help of the formula :—

Degrees of Freedom (N – 1)					
20	·687	1·72	2·09	2·53	2·84
21	·686	1·72	2·08	2·52	2·83
22	·685	1·72	2·07	2·51	2·82
23	·685	1·71	2·07	2·50	2·81
24	·685	1·71	2·06	2·49	2·80
25	·684	1·71	2·06	2·48	2·79
26	·684	1·71	2·06	2·48	2·78
27	·684	1·70	2·05	2·47	2·77
28	·683	1·70	2·05	2·47	2·76
29	·683	1·70	2·04	2·46	2·76
30	·683	1·70	2·04	2·46	2·75
35	·682	1·69	2·03	2·44	2·72
40	·681	1·68	2·02	2·42	2·71
45	·680	1·68	2·02	2·41	2·69
50	·679	1·69	2·01	2·40	2·68
60	·678	1·67	2·00	2·39	2·66
70	·678	1·67	2·00	2·38	2·65
80	·677	1·66	1·99	2·38	2·64
90	·677	1·66	1·99	2·37	2·63
100	·677	1·66	1·98	2·36	2·63
125	·676	1·66	1·98	2·35	2·62
150	·676	1·66	1·98	2·35	2·61
200	·675	1·65	1·97	2·35	2·60
300	·675	1·65	1·97	2·34	2·59
400	·675	1·65	1·97	2·34	2·59
500	·674	1·65	1·96	2·33	2·59
1000	·674	1·65	1·96	2·33	2·58
∞	·674	1·65	1·96	2·33	2·59

Example 2. Given $\text{M} = 150$

$$\sigma = 5{\cdot}4$$

$$\text{N} = 10$$

Determine the accuracy limit of 'this' M at the ·05 level and at the ·01 level.

Solution. For obtaining SE we will use the formula

$$\sigma_M = \text{SE} = \frac{\sigma}{\sqrt{\text{N}-1}} = \frac{5{\cdot}4}{\sqrt{10-1}} = \frac{5{\cdot}4}{3} = 1{\cdot}8$$

Here we see the degrees of freedom $= 10 - 1 = 9$

From the above table i.e. 'table of t' we read $t = 2{\cdot}26$ (at the ·05 level and when degrees of freedom $= 9$) and $t = 3{\cdot}25$ (at the ·01 level and when degrees of freedom $= 9$). This 't' has the same meaning which $\frac{x}{\sigma_M}$ had in table (9) i.e. it is the distance from the mean expressed in terms of the standard error of the mean. So in the above case (when degrees of freedom $= 9$) we see 95% of the sampling distribution fall between the mean and $\pm 2{\cdot}26\ \sigma_M$ and only 5% cases outside these limits. Also 99% cases fall between the mean and $\pm 3{\cdot}25\ \sigma_M$ and 1% fall outside these limits. Thus in the former case there is probability of ·05 i.e. odds are 95 : 5 or 19 : 1 that our sample mean 150 does not differ from the true mean by more than $(\pm 2{\cdot}25 \times 1{\cdot}8)$ or $\pm 4{\cdot}068$ and in the latter case there is a probability of ·99, i.e., odds are 99 : 1, that our sample mean does not differ from the true mean by more than $(\pm 3{\cdot}25 \times 1{\cdot}8)$ or $\pm 5{\cdot}85$.

If we use the formula for σ_M as

$$\sigma_M = \frac{\sigma}{\sqrt{\text{N}}}$$ for the same example we obtain

$$\sigma_M = \frac{5{\cdot}4}{\sqrt{10}} = 1{\cdot}6$$

Thus the value of σ_M is 1·6 instead of 1·8. This value of σ_M is about 11% less than the value of σ_M obtained by using another formula.

Consulting table (9) the limits of accuracy for ·05 and ·01 level would be $\pm 1{\cdot}96\ \sigma_M$ and $\pm 2{\cdot}58\ \sigma_M$ instead of $\pm\ 2{\cdot}26\ \sigma_M$ and $\pm 3{\cdot}25\ \sigma_M$ respectively. The former limits are 15% and 20% lesser than the latter limits. Thus we conclude that the sample mean in the case when N is small appears to be more reliable than it actually is when calculated by using the formula :—

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

Comparing table (9) and table of '*t*' we find that as N increases the entires of both the tables approach more closely to one another.

Say in case of table (9) 5% cases lie beyond $\pm 1{\cdot}96\ \sigma_M$ and 1% cases lie beyond $\pm 2{\cdot}58\ \sigma_M$ but in table '*t*' for $(N-1)=50$, the corresponding limits i. e. limits of ·05 level and ·01 level are $\pm 2{\cdot}01$ and $\pm 2{\cdot}68$.

When $(N-1)=1000$ these '*t*' limits are $\pm 1{\cdot}96$ and $\pm 2{\cdot}58$. Thus we can say that the 'Students' distribution becomes a normal distribution when N is large. We can use Student's ditribution both in case of small sample and large sample.

The Probable error of the mean (PE_M)

The error in the estimates may be caused due to 'chance' on account of which it appears to be unreliable. This chance error is called the 'Probable error' of the mean or PEM. Probable error is also denoted by the name Quartile deviation. Their numerical values are similar. The probable error of a calculated mean is the quartile deviation of all calculated means from the true mean and PE of a calculated coefficient of correlation is the quartile deviation of all calculated coefficients of correlation from the true coefficient.

But these terms PE and Q should be used in different ways. If we are to calculate the spread of the measures from the large number of measures we talk of the term Q. When we are estimating from a single sample, what the spread would be, we talk of the PE. In other words a Quartile deviation which is not actually measured but only estimated is called a probable error. If we consult table (9) and table (10), we find the relation between σ and PE as follows :

$$\mathrm{PE} = \cdot 6745\ \sigma$$
$$\sigma = 1\cdot 4826\ \mathrm{PE}$$

The reliability of the mean may also be determined by using $\mathrm{PE_M}$ which is obtained by multiplying σ_M by ·6745 because of the above relations between σ and PE.

$$\mathrm{PE_M} = \cdot 6745\ \sigma_M = \frac{\cdot 6745\ \sigma}{\sqrt{N}}$$

(the probable error of the mean when N is large.)

The Reliability of the median

The standard error of the median (σ_{Mdn}) is 1·2533 (approximately $\frac{5}{4}$) times the σ_M and similarly the probable error of the median (PE_{Mdn}) is 1·2533 times the PE_M .

$$\sigma_{Mdn} = \frac{1\cdot 2533\ \sigma}{\sqrt{N}} \text{ when N is large,}$$

$$\mathrm{PE}_{Mdn} = \frac{1\cdot 2533 \times \cdot 6745\sigma}{\sqrt{N}} = \frac{\cdot 8454\ \sigma}{\sqrt{N}}$$

We can also write it as

$$\mathrm{PE}_{Mdn} = \frac{1\cdot 2533\ Q}{\sqrt{N}} \text{ when N is large.}$$

But one thing should be noted that in this formula Q is the quartile deviation calculated from a frequency distribution and not the PE. We shall illustrate the method for determining the reliability of the median by using the following example.

Example 3. Given $Mdn = 72\cdot 40$

$Q = 12\cdot 84$

$N = 81$

Determine the accuracy limits of median at the ·05 level and at the ·01 level.

Solution.

$$\text{PE}_{Mdn} = \frac{1{\cdot}2533}{\sqrt{\text{N}}} \times \text{Q} = \frac{1{\cdot}2533 \times 12{\cdot}84}{9 \qquad 6}$$

$$= \frac{16{\cdot}092}{9} = 1{\cdot}788.$$

In a normal distribution the middle 95%, cases lie between the mean and ±2·9 PE, and the middle 99%, cases lie between the mean and ±3·8 PE. We may say with sufficient confidence that there is a probability of ·95 (odds are 19 : 1) that the sample Median (72·40) does not differ from the true Median by more than ± (2·9 × 1·788) or ±5·1852, i.e., the true median lies between (72·40±5·1852) or between 67·214 and 77·5852.

Again the probability is ·99 (odds are 99 : 1) that the sample median 72·40 does not differ from the true median by more than ±(3·8 × 1·788) or ±6·7944, i.e., the true median lies between (72·40 ± 6·7944) or between 65·6056 and 79·1944.

The Reliability of Measures of Variability

The Reliability of the Standard Deviation (σ)—As in the case of mean and median we calculate the corresponding standard error for determining the reliability, similarly in the case of standard deviation we firstly calculate the standard error of the standard deviation which is written as given below :

$$\left(\sigma_{\sigma} \text{ or } \text{SE}_{\sigma} \right).$$

This standard error of the standard deviation, when N is large, is given by the following formula—

$$\text{SE}_{\sigma} \text{ or } \sigma_{\sigma} = \frac{\sigma}{\sqrt{2\text{N}}}$$

when N is small say about 50 or less we should use the formula

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2(\text{N}-1)}} \text{ or } \frac{\text{S}}{\sqrt{2\text{N}}}$$

The following example will illustrate the method for determining the reliability of the standard deviation.

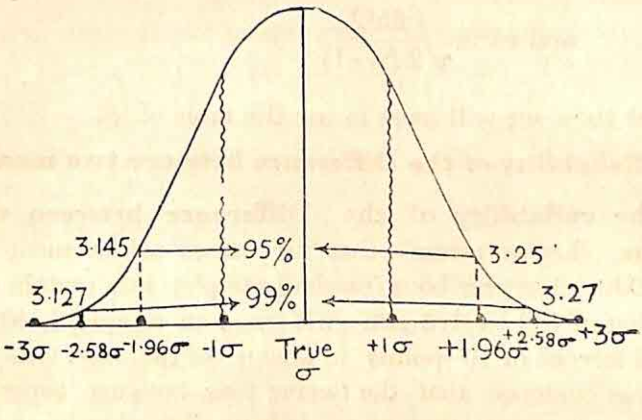
4

Example 4. Let us take the first example of this chapter in which $M=65\cdot3$, $N=6400$ and $\sigma=3\cdot2$. Let us calculate the reliability of this σ.

Solution. Here we see the sample is large and therefore

$$\sigma_\sigma = \frac{3\cdot2}{\sqrt{2\times6400}} = \frac{3\cdot2}{113\cdot14} = \cdot028$$

The probability is $\cdot95$ that the true σ is not greater than $(3\cdot2+1\cdot96\times\cdot028)$ or $3\cdot25$ and not less than $(3\cdot2-1\cdot96\times\cdot028)$ or $3\cdot145$. We can say with much confidence (Probability is $\cdot99$) that the true σ is not greater than $(3\cdot2+2\cdot58\times\cdot028)$ or $3\cdot27$ and not less than $(3\cdot2-2\cdot58\times\cdot028)$ or $3\cdot127$. This can be shown as in the following figure. Here true σ is the mean of the sampling distribution of σ 's. The σ of this normal curve is $\cdot028$ which is the standard error of the σ.



Curve 19

Normal curve showing the ·05 and ·01 levels of accuracy limits of the σ

The reliability of the Quartile Deviation (Q). Standard error of Q in terms of the distribution

$$= \sigma_Q = \frac{1\cdot11\sigma}{\sqrt{2N}}$$

Standard error of Q in terms of Q of the distribution

$$=\sigma_Q=\frac{1{\cdot}65Q}{\sqrt{2N}}.$$

Example 5. Given Mdn $=72{\cdot}40$; $Q=12{\cdot}84$; $N=81$

Determine the accuracy limits of Q at ·05 level.

Solution. $\sigma_Q=\frac{1{\cdot}65\times 12{\cdot}84}{\sqrt{2\times 81}}=1{\cdot}66$

There is a probability of ·95 that the true Q lies between $(12{\cdot}84\pm 1{\cdot}96\times 1{\cdot}66)$ or 9·59 and 16·09.

When N is small, i.e., about 50 then we should use the following formula for determining σ_Q

$$\sigma_Q=\frac{1{\cdot}11\sigma}{\sqrt{2(N-1)}} \text{ or } \frac{1{\cdot}11S}{\sqrt{2N}}$$

$$\text{and } \sigma Q=\frac{1{\cdot}65Q}{\sqrt{2(N-1)}}$$

For these we will have to use the table of '*t*'.

The Reliability of the Difference between two measures

The reliability of the Difference between the two means. Let us assume that the mean achievement score of the 400 twelve year boys (random sample) in a certain test is 50 and that of 400 twelve year girls (random sample) is 40. There is a difference of 10 points in favour of the boys which can be taken as evidence that the twelve year boys are superior than twelve year girls in that test. It is possible that if we may take various other samples of twelve year boys and girls and determine their mean achievement scores, in some cases the difference of means may be zero or in some cases may be in the favour of girls.

Our purpose is to know whether this ten points difference in favour of the boys is 'real' or only by 'chance'. We can only determine the accuracy limits of this difference within which we can say with confidence that the difference is reliable.

We say the difference to be 'significant' or 'reliable' when we are confident that this is not due to sampling fluctuations. On the other hand a difference is 'non-significant' when we are convinced that the difference may be due to sampling fluctuations.

The standard Error of the Difference when means are uncorrelated (σ_D)

The standard error of the difference between two sample means (uncorrelated) is obtained from the following formula

$$\sigma_D \text{ or } \sigma_{M_1-M_2}=\sqrt{\sigma^2_{M_1}+\sigma^2_{M_2}}$$

Where σ_D = standard error of the difference between the means which are not correlated.

σ_{M_1} = Standard error of the mean of the first group.

σ_{M_2} = Standard error of the mean of the second group.

When we take the different groups or administer different tests to the same group and get the means we say the means are unrelated.

Example 6. Two groups A and B of high school boys were tested for M and σ on an achievement test. The records obtained are as follows :—

	Number of cases	Mean	σ
Group A	625	15·8	2·5
Group B	400	15·6	2·2

Is the difference between these two means significant ?

Solution. For group A $\sigma_{M_1}=\frac{2\cdot5}{\sqrt{625}}=\frac{2\cdot5}{25}=\cdot1$

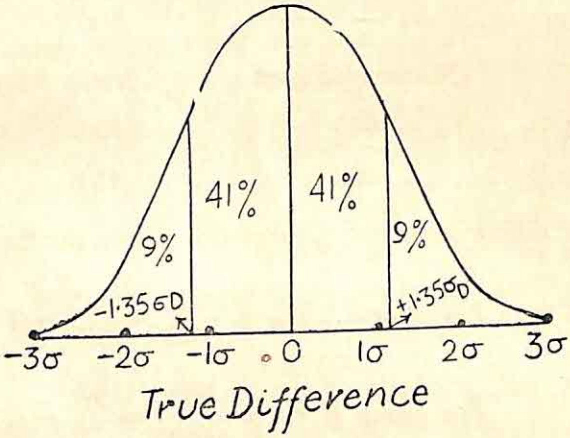
For group B $\sigma_{M_2}=\frac{2\cdot2}{\sqrt{400}}=\frac{2\cdot2}{20}=\cdot11.$

$$\therefore \sigma_D=\sqrt{\sigma^2_{M_1}+\sigma^2_{M_2}}=\sqrt{\cdot1^2+\cdot11^2}=\sqrt{\cdot0221}$$
$$=.148.$$

The actual difference between the means of the group A and group B is $\cdot20$ and σ_D of this difference is $\cdot148$. Let us assume that the difference between the population means of group A and group B is zero. Thus in all samples the difference of means of these groups will be zero except in a few cases where there can be error due to 'chance'.

This assumption we are making on the 'Null hypothesis'. Our purpose is to find out whether this difference of $\cdot20$ between the means of group A and group B is large enough to disprove the Null hypothesis. For this we firstly find out the Critical Ratio (CR).

The 'Critical Ratio' is obtained by dividing the actual difference of the means by the standard error of the difference between the means, i.e., $CR = \frac{D}{\sigma D}$ where D denotes the actual difference between the means. In our case $CR = \frac{\cdot20}{\cdot148} = 1\cdot35$ When N is large the sampling differences between the means fall



Curve 20

$$\sigma_D = \cdot148 \qquad CR = \frac{D}{\sigma_D} = 1\cdot35\sigma_D$$

in a normal curve. So we may construct a normal curve with the difference zero (the mean of our normal curve) and σ_D, i.e., 1·35 as the σ of the normal curve. The difference ·20 falls at a point 1·35 σ_D from the hypothetical mean (zero) and difference $-$·20 falls at $-$1·35 σ_D. This difference $-$·20 is obtained when the mean of the group B is higher than the mean of the group A. From table (9) we see 82% cases lie between the mean and $\pm$ 1·35 σ_D and 18% cases outside these limits. Even when we have assumed that the true difference to be zero we can expect differences greater than ·2 to occur by chance 21 times in 100 trials. This difference of $\pm$ ·2 is surely due to sampling errors. So we approve the Null hypothesis and say with confidence that the group A is not superior to group B ; the difference is only due to sampling errors. The difference is clearly not significant.

The ·05 level of significance. From table (9) we see 95% cases lie within the mean and $\pm$ 1·96 σ and 5% cases outside these limits. If a CR is greater than 1·96 and N is large, we reject the Null hypothesis with some confidence that the difference is not due to sampling fluctuations but it is significant. In our problem CR = 1·35, i.e., less than 1·96 and therefore it falls below ·05 level of significance and hence it is 'not significant'. Significance levels may also be used when we are interested in the probability that one group is better than the other. From table of '*t*' we know that 10% (P) of the cases in a normal distribution lie to the left and right of '*t*' = 1·65 ; hence 5% ($P/_2$) lie to the right of 1·65. If a CR is 1·65, we can say with confidence (on the assumption of a true difference of zero) only once in 20 trials would be a large positive difference than that obtained appear by chance.

The ·01 level of significance. From table (9) we see 99% cases lie between the mean and $\pm$ 2·58 σ and 1% cases outside these limits. If CR is greater than 2·58 we may say with more confidence that the difference is significant and not due to sampling errors. In our problem CR = 1·35 which is much less than 2·58 and hence it is not significant. If the CR is 2·33 (P = ·02 and $P/_2$ = ·01) we may be very confident odds (99 : 1) that the group now ahead is really superior to the second group in achievement.

Now we can interpret the above problem, for the significance of a difference, in another way. The difference between the means of Group A and Group B is ·2 with a σ_D of ·148. We may be confident therefore (odds are 19 : 1) that the difference between the group A and B lies within the limits ·2 $\pm(1{\cdot}96 \times {\cdot}148)$, i.e., $-{\cdot}09$ and ·49. Since the lower limit of this range is negative it is quite clear that the difference between these groups could well be zero.

Example 7. In a group intelligence test the data of two groups A and B were obtained as follows :

	N	M	σ
Group A	66	129·47	16·42
Group B	198	109·97	18·90

Is this difference significant ?

Solution. The difference between the two averages is 19·50

$$\sigma_D=\sqrt{\frac{16{\cdot}42^2}{66}+\frac{18{\cdot}90^2}{198}}=\sqrt{4{\cdot}088+1{\cdot}804}=2{\cdot}427$$

Hence the critical ratio is $\frac{19{\cdot}5}{2{\cdot}427}$ or approximately 8. Such a difference could not possibly occur through chance error of sampling. Hence the difference is significant. We can say with confidence that Group A is superior to Group B in intelligence.

The Reliability of the Difference between means in Small Independent Samples

When N's are small and the samples are independent we use the formula as $\sigma_M=\frac{S}{\sqrt{N}}=\frac{\sigma}{\sqrt{N-1}}$ for determining the standard error of the mean. In the case we should use the table of '*t*' for testing the significance of the critical ratio or '*t*'.

Example 8. In an examination the scores of 5 boys of class I were 12, 15, 10, 8 and 5 respectively. In the same examination the scores of 7 boys of class II were 7, 12, 9, 6, 13, 8, 8. Is class I significantly better than class II ?

Solution.

Class I			Class II		
Scores.	x	x^2	Scores	x	x^2
12	2	4	7	-2	4
15	5	25	12	3	9
10	0	0	9	0	0
8	-2	4	6	-3	9
5	-5	25	13	4	16
		$\Sigma x^2=58$	8	-1	1
			8	-1	1
					$\Sigma x^2=40$

5 | 50

$M_1=10$

7 | 63

$$M_2=9$$

$$(N_1-1)=5-1=4$$

$$(N_2-1)=7-1=6$$

Total degrees of freedom $=4+6=10$

$$\text{S. D. or S}=\sqrt{\frac{40+58}{10}}=\sqrt{\frac{98}{10}}=3{\cdot}13$$

$$SE_D \quad =3{\cdot}13\sqrt{\frac{12}{35}}=3{\cdot}13\times\frac{3{\cdot}46}{5{\cdot}91}=1{\cdot}8$$

$$t=\frac{10-9}{1{\cdot}8}=\frac{1}{1{\cdot}8}={\cdot}55$$

The mean of boys of class I $=10$
The mean of boys of class II$=9$
and therefore the difference between their means$=1$

When two samples are small the S. D. or S is calculated by pooling the sum of squares from two groups. We should use the formula—

$$\text{S or SD}=\sqrt{\frac{\Sigma(X_1-M_1)^2+\Sigma(X_2-M_2)^2}{(N_1-1)+(N_2-1)}}$$

In the above calculation for class I

$$\Sigma(X_1-M_1)^2=\Sigma x^2=58$$

and $$\Sigma(X_2-M_2)=\Sigma x^2=40$$

∴ Degrees of freedom for class I $=N_1-1=4$

and Degrees of freedom for class II $=N_2-1=6$

∴ Total degrees of freedom $=4+6=10$

∴ $S=\sqrt{\frac{40+58}{10}}=3{\cdot}13$. This is the S. D. of each group.

The formula for standard error of the difference between means of small independent samples is

$$SE_D \quad \text{or } S_\sigma=S\sqrt{\frac{N_1+N_2}{N_1N_2}}$$

Thus in the above problem

$$S_\sigma=3{\cdot}13\sqrt{\frac{5+7}{5\times7}}=1{\cdot}8$$

The CR or $t=\frac{D}{SE_\sigma}$ or $\frac{1}{1{\cdot}8}={\cdot}55$

Our purpose is to test the significance of the mean difference. From the table of 't' for 10 degrees of freedom we find the entry 1·81 at the ·01 level and 2·23 at ·05 level. The critical ratio ·55 falls below these values and hence the difference is not significant at the ·05 level and ·01 level. Thus we have no evidence in believing that class I is superior to class II. We should note that at ·01, 5% of our t's (CR's) lie to the right of +1·76 and 5% left of 1·76. The limit at ·01 (not at ·05) must be taken therefore, to give the ·05 significance level, if we are interested in knowing the probability that 'the given difference or a greater positive one' may arise from sampling errors.

Example 9. The mean score of 41 boys in a reading test was 43·8 with $\sigma=6{\cdot}3$ and the mean score of 31 girls on the same test was 40·5 with $\sigma=6{\cdot}1$. Is the mean difference of 3·3 between the boys and girls significant ?

Solution. $\sigma_D=\sqrt{\sigma_{M1}^2+\sigma_{M2}^2}=\sqrt{\frac{(6{\cdot}3)^2}{40}+\frac{(6{\cdot}1)^2}{30}}$

$=1{\cdot}5$

Here the standard errors of mean are calculated by using the formula $\sigma_M = \frac{\sigma}{\sqrt{N-1}}$

The actual difference between the means $= 3{\cdot}3$

Hence critical ratio or $t = \frac{3{\cdot}3}{1{\cdot}5} = 2{\cdot}2$

$$
\begin{aligned}
\text{The degrees of freedom} &= (N_1-1)+(N_2-1)\\
&= (41-1)+(31-1)\\
&= 70
\end{aligned}
$$

When the degrees of freedom $= 70$ a 't' of $\pm$ 2·00 or more may be expected on the Null hypothesis 5% of the time and a 't' of $\pm$ 2·65 or more may be expected 1% of the time. The obtained t of 2·2 passes the ·05 but not the ·01 level. We may therefore reject the Null hypothesis with considerable confidence. We can conclude that the (odds 19 : 1) mean difference lies between 3 and 6·3 ($3{\cdot}3 \pm 2{\cdot}0 \times 1{\cdot}5$).

The Standard Error of the difference between two means, when means are correlated

(a) Single Group Method. This case arises when the same test is administered to the same group but at different intervals. Say a group of boys have been tested today in a certain mechanical arithmetic test. If after a few months we administer the same test to the same group to measure the effect of practice, we may find certain differences in the scores of the boys. Our purpose is to determine whether this difference is significant or not. The formula for the standard error of the mean when the means are correlated is

$$\sigma_D = \sqrt{\sigma^2_{M1} + \sigma^2_{M2} - 2r_{12}\sigma_{M1}\sigma_{M2}}$$

Where σ_{M1} and σ_{M2} are standard errors of the mean of the pretest and the post test and r_{12} is the coefficient of correlation between the scores of the pretest and the post-test.

Example 10. 82 boys were tested in the beginning and in the end of the year on the same test. Their mean in the pretest was 40 with a σ_1 of 5 and the mean in the post test was

45 with a σ_2 of 4. The coefficient of correlation between the scores on the pretest and the post test was ·7. Is the difference between the mean scores on the pretest and the post-test significant to tell upon the progress of the boys.

Solution. $\sigma_{M1}=\frac{\sigma_1}{\sqrt{N_1-1}}=\frac{5}{\sqrt{81}}=\frac{5}{9}=\cdot 55$

$$\sigma_{M2}=\frac{\sigma_2}{\sqrt{N_2-1}}=\frac{4}{\sqrt{81}}=\frac{4}{9}=\cdot 44$$

$$r_{12}=\cdot 7$$

$$\therefore \quad \sigma_D=\sqrt{(\cdot 55)^2+(\cdot 44)^2-2\times\cdot 7\times\cdot 44\times\cdot 55}$$

$$=\cdot 396$$

Here we see the degrees of freedom$=82-1=81$ because 82 boys will have 82 scores in the pretest and 82 scores in the post test and therefore there will be 82 differences

$$\text{CR or } t=\frac{D}{\sigma_D}=\frac{5}{\cdot 396}=12\cdot 6$$

The t for 81 degrees of freedom is 2·38 at the ·02 level. Our t is far greater than this 2·38. Therefore the gain in the means is significant. There is a probability much less than ·01 that the gain in the mean may be due to sampling fluctuations.

We can say with much confidence that the boys progressed well during the year. We should note it carefully that the entries from the ·01 and ·02 columns are taken for significance levels ·05 and ·01 because we are dealing with the probability of a gain (Positive difference).

(**b**) **Equivalent Group Method.** If the same test be administered to two different groups of same educational standard and after this pretest the former group be encouraged or praised in some way or the other but not the latter group, it is possible that the former group may do better in the post test than the latter group on account of this incentive. Our purpose is to determine whether this gain in performance or attainment is significant or not. The former is called the 'experimental group' and the latter the 'control group.'

The standard error of the difference between the means of equivalent groups is found from the same formula which was used in the case of single group method. Here r_{12} is the coefficient of correlation between final scores on the post-test of two groups who are matched in the pretest.

Example 11. Two groups of high school pupils are matched for initial ability in a biology test. Group I is taught by the lecture method and group II by the lecture demonstration method. Data are as follows :

	Group I (Control)	Group II (Experimental)
N	60	60
Mean initial score on the biology test	42·30	42·50
σ of initial scores on the biology test	5·36	5·38
Mean final score on the biology test	54·54	56·74
σ of final scores on the biology test	6·34	7·25
r_{12} (between final scores on biology test)	=·50	

Is the difference between the final scores made by group I and II upon the biology test significant at the ·05 level ? at the ·01 level ?

Solution. We see the mean initial scores of both the groups are almost the same.

The difference between the means of the final test

$$=(56{\cdot}74-54{\cdot}54)=2{\cdot}2$$

$$\sigma_{M1}=\frac{6{\cdot}34}{\sqrt{60-1}}=\frac{6{\cdot}34}{\sqrt{59}}={\cdot}82$$

$$\sigma_{M2}=\frac{7{\cdot}25}{\sqrt{60-1}}=\frac{7{\cdot}25}{\sqrt{59}}={\cdot}94$$

$$\sigma_D=\sqrt{({\cdot}82)^2+({\cdot}94)^2-2\times{\cdot}5\times{\cdot}82\times{\cdot}94}$$
$$={\cdot}8861$$

$$\text{Critical ratio or } t = \frac{2{\cdot}2}{{\cdot}8861} = 2{\cdot}47$$

There are 59 degrees of freedom. From the table of 't' we can say that the group II is significantly superior to group I (control) at the $\frac{P}{2} = {\cdot}05$ level ($t = 1{\cdot}67$) and also significantly superior at $\frac{P}{2} = {\cdot}01$ level ($t = 2{\cdot}39$). While consulting the table we may take t's for 60 degrees of freedom instead of 59 because they are quite close.

(c) **Matched group.** The groups are matched when they are made alike as regards the mean and SD in some measure. It is not necessary to pair off the individuals as in the case of equivalent group and the N's of two groups are not necessarily to be equal, but there should not be a large difference between the N's of the two groups. Suppose we are to study the mechanical ability of the two groups (say in x test), and the groups have already been matched for M and S. D. in general intelligence (say in y test, which is a variable,) then the standard error of the difference between means in x test is

$$SE_{D_{M_1 - M_2}} = \sigma_D = \sqrt{\left(\sigma^2_{MX_1} + \sigma^2_{MX_2}\right)\left(1 - r^2_{xy}\right)}$$

Where r_{xy} is correlation coefficient between x and y in the population from which our matched samples are drawn.

Example 12. Two groups of high-school students are matched for M and σ upon a group intelligence test. There are fifty eight students in Group A and seventy two in Group B. The records of these two groups upon a battery of 'learning' tests are as follows :

	Group A	Group B
M	48·52	53·61
σ	10·60	15·35
N	58	72

The correlation of the group intelligence test and the learning battery in the entire group from which A and B were drawn is ·50. Is the difference between Groups A and B significant at the ·05 level? at the ·01 level?

Solution. $Mx_1 - Mx_2 = 53{\cdot}61 - 48{\cdot}52 = 5{\cdot}09$

$$\sigma_D = \sqrt{\left[\frac{(10{\cdot}6)^2}{58-1} + \frac{(15{\cdot}35)^2}{72-1}\right]\left[1 - {\cdot}5^2\right]}$$

$$= \sqrt{\left[\frac{(10{\cdot}6)^2}{57} + \frac{(15{\cdot}35)^2}{71}\right]\left[\frac{3}{4}\right]}$$

$$= 1{\cdot}99$$

$$t \text{ or CR} = \frac{D}{\sigma_D} = \frac{5{\cdot}09}{1{\cdot}99} = 2{\cdot}56$$

Degrees of freedom $= 57 + 71 - 1 = 127$.

Since $t = 2{\cdot}56$, hence from table of 't' we find that the difference is significant at $\frac{P}{2} = {\cdot}05$ level ($t = 1{\cdot}66$) and also significant at the $\frac{P}{2} = {\cdot}01$ level ($t = 2{\cdot}36$).

We can say with confidence that the group B is superior to Group A in learning though both are equal in intelligence.

The Reliability of the Difference between Medians

The formula for the standard error of the difference between uncorrelated medians is as follows :

$$\sigma_D = \sigma_{Mdn1 - Mdn2} = \sqrt{\sigma^2_{Mdn1} + \sigma^2_{Mdn2}}$$

The reliability of the difference between two medians can be computed in the same way as the reliability of the difference between the two means.

The Reliability of the Difference between σs

(1) The SE of a difference between σs which are uncorrelated. The difference in σ can be due to many factors, as in the case of a teacher if he adopts a new method of teaching instead of old method there may be a certain change in variability.

The standard error of the difference between uncorrelated σ's is given by

$$\sigma_{D_{\sigma}} \text{ or } \sigma_{(\sigma_1-\sigma_2)}=\sqrt{\sigma^2_{\sigma_1}+\sigma^2_{\sigma_2}}$$

Where $\sigma_{D_{\sigma}}$ = Standard error of the difference between uncorrelated σ's

σ_{σ_1} = standard error of the first σ.

σ_{σ_2} = standard error of the second σ.

Example 13. Two groups A and B of high school boys were tested for M and σ on an achievement test. The records obtained are as follows :

	N	M	σ
Group A	625	15·8	2·5
Group B	400	15·6	2·2

Is the difference between two σ's singnificant ?

Solution. $$\sigma_{\sigma_1}=\frac{2{\cdot}5}{\sqrt{2\times625}}={\cdot}070$$

$$\sigma_{\sigma_2}=\frac{2{\cdot}2}{\sqrt{2\times400}}={\cdot}078$$

$$\sigma_{D_{\sigma}}=\sqrt{({\cdot}070)^2+({\cdot}078)^2}$$

$$={\cdot}1047$$

The obtained difference between σs = 2·5 − 2·2 = ·3

CR or $t=\frac{{\cdot}3}{{\cdot}1047}=2{\cdot}86$.

From table we find that our $t=2{\cdot}86$ is greater than 1·96 (·05 level of significance) and also than 2·58 (·01 level of significance) and hence we reject the Null hypothesis.

We can say with considerable confidence that the difference between σs is significant and not due to sampling fluctuations.

(2) The standard error of a difference between σ's which are correlated. When the SDs of the same group upon two occasions are compared or the SD's of equivalent groups on a final test are compared, we must take into account the correlation between the SD's of the groups being compared.

The following formula is used for finding out the standard error of the difference between correlated σs.

$$\sigma_{D_\sigma}=\sqrt{\sigma^2_{\sigma_1}+\sigma^2_{\sigma_2}-2r^2_{12}\,\sigma_1\,\sigma_2}$$

Where r_{12} is the coefficient of correlation between scores in initial and final tests of the same group.

Example 14. The same as in Example (**11**).

Here we may find whether the difference in the variability of the final scores made by groups I and II is significant at ·05 level.

Solution. $\sigma_{\sigma_1}=\dfrac{6{\cdot}34}{\sqrt{2\times59}}={\cdot}583$

$$\sigma_{\sigma_2}=\frac{7{\cdot}25}{\sqrt{2\times59}}={\cdot}667$$

$$r_{12}={\cdot}5$$

Obtained difference between σ's = ·91

$$\therefore\ \sigma_{D_\sigma}=\sqrt{(5{\cdot}83)^2+({\cdot}667)^2-2\times{\cdot}5^2\times{\cdot}583\times{\cdot}667}$$

$$={\cdot}7683$$

$$t \text{ or CR}=\frac{{\cdot}91}{{\cdot}7683}=1{\cdot}18$$

From table of '*t*' for 59 degrees of freedom this '*t*' is not significant at the P = ·05 level (2·00) nor in the positive direction at the P/2 = ·05 level (1·67). There is no evidence, then, that the lecture demonstration method increased the variability.

The Null hypothesis

According to the null hypothesis, when we draw random samples from the same population no true difference exists between the samples. If at all it exists, it is due to sampling fluctuations or errors. If we may be able to say with considerable confidence that the difference that occurs between the two samples is significant we reject the null hypothesis. The null hypothesis is retained when we have no sufficient evidence to prove the existence of the difference between the two samples, i.e., when the difference is not significant. We should note it carefully that the disapproval of null hypothesis does not mean the acceptance of a contrary hypothesis. If the same achievement test in reading be administered to a sample of twelve year boys and another sample of twelve year girls and if the difference between their means be five points in favour of boys we can find out whether this difference is significant or not. Say this difference is significant, still then we cannot say that the boys are better in that ability. We can only be sure of a significant difference between the mean scores of boys and girls. In subsequent comparisons of boys and girls, if all experimental variables likely to influence the reading score are controlled and the difference still remains, we may then be willing to assert the existence of a true sex difference in that ability. But the acceptance of null hypothesis usually ends in a definite judgment.

Testing of the null hypothesis against the direct determination of the probable outcomes

Suppose we toss a coin. In doing so the probability of getting the 'head' is $\frac{1}{2}$ and the probability of getting the 'tail' is $\frac{1}{2}$. If we toss the coin any number of times each time we will have the same probability ($\frac{1}{2}$) for getting the head and $\frac{1}{2}$ for getting the tail. These are the expected results. But we may get 7 heads and 3 tails in ten tosses of the coin. These are experimentally observed results. The null hypothesis can be tested by comparing the expected and observed results.

Example 15. Two sharp clicking sounds are presented in succession, the second being always more intense or less intense than the first. Presentation is in random order. In eight trials an observer is right six times. Is the result significant ?

Solution. We will test the result against the binomial expansion because the judgments are separate and independent. The eight trials will have eight judgments. We can treat this question as if eight coins have been tossed and each right judgment corresponds to head and each wrong judgment corresponds to tail. As we have read before we are likely to expect 4 right judgments (heads) and four wrong judgments (tails) in our 8 trials. But we get six right judgments and two wrong judgments. Now we are to see whether six right judgments are significantly greater than four right judgments that we expected. The probability of 8 right judgments, 7 right and one wrong judgments, 6 right and two wrong judgments can be obtained from the following Binomial Expansion—

$$(H+T)^8=H^8+8H^7T+28H^6T^2+56H^5T^3+70H^4T^4$$
$$+56H^3T^5+28H^2T^6+8HT^7+T^8$$

Total number of happenings

$$=1+8+28+56+70+56+28+8+1=256$$

$$\text{Probability of 8 right judgments} = \frac{1}{256}$$

$$\text{Probability of 7 right and one wrong judgments} = \frac{8}{256}$$

$$\text{Probability of 6 right and two wrong judgments} = \frac{28}{256}$$

Thus the probability of getting six or more right judgments

$$=\frac{1}{256}+\frac{8}{256}+\frac{28}{256}=\cdot 145$$

The probability of getting just six right judgments

$$=\frac{28}{256}=\cdot 109$$

Neither of these results is significant at the ·05 level of confidence and accordingly the null hypothesis must be retained. The observers judgments cannot be believed to be better than chance expectations.

Testing the Reliability of a coefficient of correlation against the Null hypothesis

The significance of an obtained r may be tested more exactly against the null hypothesis than in terms of PE_r.

The formula

$$\sigma_r = \frac{(1-r^2)}{\sqrt{N-1}}$$

gives the standard error of a coefficient of correlation. The probable error of a coefficient of correlation is given by

$$PE_r = \frac{\cdot 6745\,(1-r^2)}{\sqrt{N-1}}.$$

PE_r is more often used. Suppose $r = \cdot 60$ and $N = 120$ then $PE = \cdot 04$ to two decimals. This probable error is taken to mean that the chances are 50 in 100 (odds 1 : 1) that the obtained r of ·60 does not miss the true value or the population value by more than ±·04.

But there are two objections to use these above formulae.

(1) r in these formulae is really the population r. Since we do not have the true r, we must substitute the calculated or sample r in the formula in order to get an estimate of the standard or probable error. If the obtained r is in error, our estimate will also be in error ; and at best it is approximate.

(2) In the second place, the sampling distribution of r is not normal except when the population r is ·00 and N is large. When r is high (·80 or more) and N is small, the sampling distribution of r is skewed and PE is decidedly misleading.

When we test it against the null hypothesis it is done more exactly. Assuming the population r to be zero, the method consists in comparing the t value (See table of 't') for the obtained r with the 't's to be expected by chance at the ·05 and ·01 level. The t for given r is found from the formula

$$t=\frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

in which r=obtained coefficient
N=number of cases.

The value of t may be read from the 'table of t' which is entered with N−2 degrees of freedom. To illustrate suppose $r=0{\cdot}9$ and $N=100$. Then from the above formula

$$t=\frac{\cdot 9\sqrt{\cdot 98}}{\sqrt{1-\cdot 9^2}}=20{\cdot}5.$$

Entering table of t for the 98 degrees of freedom ($N-2=98$) we find that t at ·05 level is 1·98 and at ·01 level 2·63. Since our t is far greater than the second of these values, we conclude forth-with that the null hypothesis is clearly disproved and our r is very significant. The probability that we should have obtained an r of ·90, if the true r were ·00, is much less than ·01.

SUMMARY

(1) (The Standard Error of the Arithmetic mean when N is large)

$$=\sigma_M=\frac{\sigma}{\sqrt{N}}$$

(2) (The Standard Error of the mean in small samples)

$$=\sigma_M=\frac{\sigma}{\sqrt{N-1}}$$

(3) (The Probable Error of the Arithmetic Mean when N is large, i e. greater than 50)

$$=PE_M=\frac{\cdot 6745\sigma}{\sqrt{N}}$$

(4) (Standard Error of the median when N is large)

$$=\frac{1{\cdot}2533\sigma}{\sqrt{N}}=\sigma_{Mdn}$$

(5) (Probable Error of the median when N is large)

$$=PE_{Mdn}=\frac{{\cdot}8454\sigma}{\sqrt{N}}$$

$$=\frac{1{\cdot}2533Q}{\sqrt{N}}$$

(6) (Standard Error of a S. D. when N is large)

$$=\sigma_{\sigma}=\frac{\sigma}{\sqrt{2N}}$$

(7) (Standard Error of a S. D. when N is small)

$$=\sigma_{\sigma}=\frac{\sigma}{\sqrt{(2N-1)}}$$

(8) (Standard Error of Q in terms of σ of the distribution)

$$=\sigma_{\sigma}=\frac{1{\cdot}11\sigma}{\sqrt{2N}}$$

(9) (Standard Error of Q in terms of Q of the distribution)

$$=\sigma_{Q}=\frac{1{\cdot}65_{Q}}{\sqrt{2N}}$$

(10) (Standard Error of the difference between two uncorrelated means when N is large)

$$=\sigma_{D}=\sigma_{(M1-M2)}=\sqrt{\sigma_{M1}+\sigma^{2}_{M2}}$$

(11) (Probable Error of the difference between two means which are uncorrelated ; N being large)

$$=PE_{D}=\sqrt{PE^{2}_{M1}+PE^{2}_{M2}}$$

(12) (Standard Deviation when two small independent samples are pooled)

$$= SD \text{ or } S = \sqrt{\frac{\Sigma(X_1 - M_1)^2 + \Sigma(X_2 - M_2)^2}{(N_1 - 1) + (N_2 - 1)}}$$

(13) (Standard Error of the difference between means in small independent samples)

$$= S_D \text{ or } SE_D = s\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

(14) (Standard Error of the difference between correlated means)

$$= \sigma_D = \sqrt{\sigma^2_{M1} + \sigma^2_{M2} - 2r_{12}\,\sigma_{M1}\,\sigma_{M2}}$$

(15) (Standard Error of the difference between M's of matched groups.)

$$= \sigma_D = \sqrt{\left(\sigma^2_{M1} + \sigma^2_{M1}\right)\left(1 - r^2_{xy}\right)}$$

(16) (Standard Error of the difference between uncorrelated σ's)

$$= \sigma_{D_\sigma} = \sqrt{\sigma^2_{\sigma_1} + \sigma^2_{\sigma_2}}$$

(17) (Standard Error of the difference between correlated σ's)

$$= \sigma_{D_\sigma} = \sqrt{\sigma^2_{\sigma_1} + \sigma^2_{\sigma_2} - 2r^2_{12}\,\sigma_{\sigma_1}\,\sigma_{\sigma_2}}$$

(18) (Standard Error of coefficient of correlation)

$$= \sigma_r = \frac{1 - r^2}{\sqrt{N - 1}}$$

(19) (Probable Error of a coefficient of correlation)

$$= PEr = \frac{\cdot 6745\,(1 - r^2)}{\sqrt{N - 1}}$$

(20) $\left(\begin{array}{c}t \text{ for determining the significance of a computed} \\ r \text{ on the null hypothesis}\end{array}\right)$

$$=t=\frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

EXERCISES

1. Given : $M=24{\cdot}60$; $\sigma=3{\cdot}20$; $N=100$

(a) Determine the accuracy limits of this mean at the ·05 level ; at the ·01 level.

(b) Determine the accuracy limits of σ at the ·05 & ·01 level.

Ans. (a) 25·77 & 27·03 ; 25·57 & 27·23.
(b) 2·75 & 3·65 ; 2·61 & ·379.

2. The difference between two means (M_1-M_2) is 3·60 and $\sigma_D=3{\cdot}00$. The samples are large.

(a) Is the obtained difference significant at the ·05 level ?

(b) What percent is the obtained difference of the difference necessary for significance at the ·01 level ?

Ans. (a) C R = 1·20
(b) 46·5%.

3. In 1883, the Anthropometric committee of the British Association found the mean height of 8585 adult males in the British Isles to be 67·46 inches, with a σ of 2·57 inches. How reliable is this mean ? How much does it probably diverge from the mean which would have been obtained had all adult males in the British Isles been measured ?

Ans. $\sigma_M=\cdot028$ inches.

(odds 19 : 1) that 67·46 inches does not differ from the true mean height by more than ± ·05 inches.

(odds 99 : 1) that 67·46 inches does not differ from the true mean height by more than ± ·07 inches.

4. On the Trabue Language Scale A, 801 twelve year old boys made the following records : Median=21·4 ; Q=4·9. How reliable is this median ? How well does it represent the median of twelve year old boys in general on the given scale ?

Ans. $PE_{Mdn} = \cdot 22$

odds are (19 : 1) that 21·4 does not miss the true value of median by more than ±·64 and odds are (99 : 1) that 24·1 does not miss the population median by more than ±·84.

5. The average scores for 136 men and 114 women students on the Cattell Intelligence Test were 101·55 and 99·65 with σ of 12·36 and 9·95 respectively. Is this difference in variability significant ?

Ans. $\sigma_{D_\sigma} = 0{\cdot}998.$

The odds against such a difference occurring through chance errors are 129 : 1. There is strong evidence that men students of the type tested are in general more spread out in their ability at an intelligence test than are women students.

6. Random samples of 18 boys and 13 girls are given a certain test.

Their scores are :—

Boys : 48, 60, 81, 19, 38, 26, 74, 23, 63, 58, 92, 71, 34, 18, 45, 64, 41, 59.

Girls : 45, 23, 39, 68, 31, 7, 63, 0, 16, 56, 49, 32, 23.

Is there significant difference between mean scores of boys and girls ?

Ans. $t = 2{\cdot}04$. The difference between the two sexes is significant.

7. Two groups X and Y, of seventh grade children are paired child for child for age and for score upon form A of the Otis group Intelligence Scale. Three weeks later, both groups are given Form B of the same test. Before the second test, Group X, the experimental group, is praised for

its performance on the first test and urged to better its score if possible. Group Y, the control group, is given the second test without comment. Will the incentive (praise) serve to increase significantly the final score of Group X over Group Y ?

Data are as follows :

	Experimental Group X	Control Group Y
No. of children in each Group	72	72
Mean scores on Form A (initial test)	80·42	80·51
S. D. on Form A (initial test)	23·61	23·45
Mean scores on Form B (final test)	88·63 (M_1)	83·24 (M_2)
S. D. on Form B (final test)	64·36 (σ_1)	21·62 (σ_2)
Gain ($M_1 - M_2$)	5·39	
Standard error of the means, final tests	2·89	2·57

Correlation between final, scores (experimental and control groups) = ·65

Ans. $t = 2{\cdot}34$, $df = 71$.

The incentive group is significantly superior to the control group at ·05 level and almost so at ·01 level.

8. The achievement of two groups of first-year high school boys, the one from an academic, the other from a technical high school, is compared upon a Mechanical Ability test. The two groups are matched for mean and σ upon a general intelligence test so that the experiment becomes one of comparing the mechanical ability scores of two groups of boys of 'equal' general intelligence enrolled in different curricula. Data are as follows :

	Academic	Technical
No. of boys in each group :	125	137
Means on intelligence test (Y)	102·50	102·80
σs on intelligence test (Y)	33·55	31·62
Means on Mechanical Ability test (X)	51·42	54·38
σ's on Mechanical Ability test (X)	6·24	7·14

Correlation between the General Intelligence test and Mechanical Ability test for first year high school boys is 30.

Is the difference between two groups significant ?

Ans. $t = 3{\cdot}75$; Mean difference is highly significant.

9. An Intelligence Test of 11+was given to 936 twins. The mean score was found 32·02 and S. D. to be equal 16·46. The same test was repeated on the entire population of 70805 including twins. In this case mean was 36·741 and S. D. was 16·10. Find whether the difference between the two standard deviations is significant.

Ans. $^{\sigma_D}\sigma = \cdot 382$

$t = \cdot 94$

The difference between the two σs is not significant at ·05 level or ·01 level.

10. The following is the data for the general science test scores of rural and urban pupils reading in the eleventh class. Is the difference between their means significant ?

	Number	Mean	S. D.
Rural Pupils	56	39·9	11·8
Urban Pupils	104	40·56	10·6

Ans. C R = ·308

The difference is neither significant at ·01 level nor at ·05 level.

CHAPTER 24

THE χ^2 (CHI-SQUARE) TEST OF SIGNIFICANCE

IN THE PREVIOUS CHAPTER WE READ ABOUT THE TESTING OF hypothesis concerning difference of various sorts, differences between means, σ's etc. We also made the use of the null hypothesis in testing the significance of the differences between the two groups.

In our present chapter we will see how χ^2 affords a measure of correspondence between fact and theory, i.e. observed and expected results. The formula for calculating χ^2 (chi-square) is

$$\text{chi-square or chi}^2 \text{ or } \chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right]$$

in which—

f_o = frequency of occurrence of observed or experimentally determined facts.

f_e = expected frequency of occurrence on some hypothesis.

χ^2 = measure of divergence between observed frequency and expected frequency.

χ^2 is obtained by squaring the difference between the observed and expected frequencies; dividing the square of the differences by the expected number in each case and then summing all the quotients.

If the observed frequency is close to the expected frequency, the value of χ^2 is small and hence there will be a close correspondence between the observed data and the hypothesis that is to be tested. If χ^2 is great we can expect a considerable divergence between the observed and expected results. To evaluate χ^2 we consult the table of 'χ^2' where '*df*' denotes the number of degrees of freedom. If our contingency table has 3 rows and 5 columns, the degrees of freedom (*df*) will be $(3-1)(5-1)=2\times 4=8$.

If there are r rows and c columns, the degrees of freedom will be $(r-1)(c-1)$. To obtain the significance of χ^2 we consult the table of χ^2 and find the P for the corresponding number of degrees of freedom.

[Testing the divergence of observed values from values calculated on the hypothesis of equal probabillity (Null hypothesis].

Example 1. 200 digits were chosen at random from a set of tables. The frequencies of the digits were :

Digit	0	1	2	3	4	5	6	7	8	9	Total
Frequency (f_o)	18	19	23	21	16	25	22	20	21	15	200

Use the χ^2 test to assess the correctness of the hypothesis that the digits were distributed in equal numbers in the table from which these were chosen.

Solution. The data can be put in the contingency table as shown below, since there are 200 digits the expected frequency (fe) in each case will be $\frac{200}{10}=20$. These expected frequencies have been entered in the second row of the contingency table.

Digit	0	1	2	3	4	5	6	7	8	9	Total
Observed frequency (f_o)	18	19	23	21	16	25	22	20	21	15	200
Expected frequency (f_e)	20	20	20	20	20	20	20	20	20	20	200

$(f_o - f_e) = 2 \quad 1 \quad 3 \quad 1 \quad 4 \quad 5 \quad 2 \quad 0 \quad 1 \quad 5$

(Difference of each f_o and f_e)

$(f_o - f_e)^2 = 4 \quad 1 \quad 9 \quad 1 \quad 16 \quad 25 \quad 4 \quad 0 \quad 1 \quad 25$

squaring each $(f_o - f_e)$

$$\frac{(f_o-f_e)^2}{f_e} = \frac{4}{20} \quad \frac{1}{20} \quad \frac{9}{20} \quad \frac{1}{20} \quad \frac{16}{20} \quad \frac{25}{20} \quad \frac{4}{20} \quad \frac{0}{20} \quad \frac{1}{20} \quad \frac{25}{20}$$

$$x^2 = \Sigma\left[\frac{(f_o-f_e)^2}{f_e}\right] = \frac{4}{20}+\frac{1}{20}+\frac{9}{20}+\frac{1}{20}+\frac{16}{20}+\frac{25}{20}+\frac{4}{20}$$
$$+\frac{0}{20}+\frac{1}{20}+\frac{25}{20}$$
$$=\frac{86}{20}=4{\cdot}3.$$

df = degrees of freedom $= (2-1)\ (10-1) = 1 \times 9 = 9$

[because there are two rows and ten columns]

Entering the χ^2 table we find in row $df=9$, a χ^2 of almost 4·3 (actually 4·168) in the column headed by ·90. This ·90 (P) is the probability. We regard ·05 level as the significant level. Again from table of χ^2 we find for 9 degrees of freedom the value of P at ·05 is 16·919. Our value of χ^2 (4·3) is much less than this (16·919) and hence we can say with much confidence that our difference is not significant. The null hypothesis is retained. We can believe that the divergence between the expected and observed frequencies is attributed to sampling fluctuations. The null hypothesis seems to be reasonable.

[**Table of** χ^2]

P gives the probability of exceeding the tabulated value of χ^2 for the specified number of degrees of freedom (*df*). The values of χ^2 are printed in the body of the table.

Adapted from R.A. Fisher's statistical method for research workers, Oliver and Boyd.

df	0·95	0·90	0·80	0·70	0·50	0 30	0·20	0·10	0·05	0·02	0·01
1	0·00393	0·0158	0·0642	0·148	0·455	1·074	1·642	2·706	3·841	5·412	6·635
2	0·103	0·211	0·446	0·713	1·386	2·408	3·219	4·605	5·991	7·824	9·210
3	0·352	0·584	1·005	1·424	2·366	3·665	4·642	6·251	7·815	9·837	11·345
4	0·711	1·064	1·649	2·195	3·357	4·878	5·989	7·779	9·488	11·668	13·277
5	1·145	1·610	2·343	3·000	4·351	6·064	7·289	9·236	11·070	13·388	15·086
6	1·635	2·204	3·070	3·828	5·348	7·231	8·558	10·645	12·592	15·033	16·812
7	2·167	2·833	3·822	4·671	6·346	8·383	9·803	12·017	14·067	16·622	18·475
8	2·733	3·490	4·594	5·527	7·344	9·524	11·030	13·362	15·507	18·168	20·090
9	3·325	4·168	5·380	6·393	8·343	10·656	12·242	14·684	16·919	19·679	21·666
10	3·940	4·865	6·179	7·267	9·342	11·781	13·442	15·987	18·807	21·161	23·209
11	4·575	5·578	6·989	8·148	10·341	12·899	14·631	17·275	19·675	22·618	24·725
12	5·226	6·304	7·807	9·034	11·340	14·011	15·812	18·549	21·026	24·054	26·217
13	5·892	7·042	8·634	9·926	12·340	15·119	16·985	19·812	22·362	25,472	27·688
14	6·571	7·790	9·467	10·821	13·339	16·222	18·151	21·064	23·685	26·873	29·141
15	7·261	8·547	10·307	11·721	14·339	7·322	19·311	22·307	24·996	28·259	30·578
16	7·962	9·312	11·152	12·624	15·338	18·418	20·465	23·542	26·296	29·633	32·000
17	8 672	10·085	12·002	13·531	16·338	19·511	21·615	24·769	27·587	30·995	33·409
18	9·390	10·865	12·857	14·440	17·338	20·601	22·760	25·989	28·869	32·346	34·805
19	10·117	11 651	13·716	15·352	18·338	21·689	23·900	27·204	30·144	33·687	36·191
20	10·851	12·443	14·578	16·266	19·337	22·775	25·038	28·412	31·410	35·020	37·566
21	11·591	13·240	15·445	17·182	20·337	23·858	26·171	29·615	32·671	36·343	38·932
22	12·338	14·041	16 314	18·101	21·337	24·939	27·301	30·813	33·924	37·659	40·289
23	13·091	14·848	17·187	19·021	22·337	26·018	28·429	32 007	35·172	38·968	41·638
24	13·848	15·659	18·062	19·943	23·337	27·096	29·553	33·196	36·415	40 270	42·980
25	14 611	16·473	18·940	20·867	24·337	28·172	30·675	34·382	37·652	41·566	44·314
26	15·379	17·292	19·820	21·792	25·336	29·246	31·795	35·563	38·885	42·856	45·642
27	16·151	18·114	20·703	22·719	26·336	30·319	32·912	36·741	40·113	44·140	46·963
28	16·928	18 939	21·588	23·647	27·336	31·391	34·027	37·916	41·337	45·419	48·278
29	17·708	19·768	22·475	24·577	28·336	32·461	35·139	39·087	42·557	46·693	49·588
30	18·498	20·599	23 364	25·508	29·336	33·530	36·250	40·256	43·773	47·962	50·892

Example 2. In 120 throws of a single die, the following distribution of the faces was obtained.

	Faces						
	(1)	(2)	(3)	(4)	(5)	(6)	Total
observed frequency	30	25	18	10	22	15	120

Do the results constitute a refutation of the "equal probability" (null) hypothesis ?

Solution. Since there are 6 faces we can expect each face to occur $\frac{120}{6}=20$ times in 120 throws. Thus we can make the following table

Faces	(1)	(2)	(3)	(4)	(5)	(6)	Total
Observed frequencies (f_o)	30	25	18	10	22	15	120
expected frequencies (f_e)	20	20	20	20	20	20	120

$(f_o - f_e) =$ 10 5 2 10 2 5

$(f_o - f_e)^2 =$ 100 25 4 100 4 25

$\frac{(f_o - f_e)^2}{f_e} =$ 5, 1·25, ·2, 5, ·2, 1·25

$x^2 = 5 + 1{\cdot}25 + {\cdot}2 + 5 + {\cdot}2 + 1{\cdot}25 = 12{\cdot}90$

Degrees of freedom $= (r-1)\ (c-1) = (2-1)\ (6-1)$
$= 1 \times 5 = 5$

We consider ·05 level as significant level. From the table χ^2 we see for $d_f = 5$ the value of P at ·05 is 11·070. Our χ^2 (12·9) is more than this (11·070) and hence the null hypothesis is refused. The divergence is significant. We can say the results constitute the refutation of the "equal probability" (null) hypothesis.

Testing Divergence of observed values from values calculated on the hypothesis of a normal Distribution

We can expect that the frequencies of an event observed may follow the normal distribution instead of being equally probable. The sample should be random. In that case the formula for calculating χ^2 remains the same but our expected frequencies are those which form a normal curve.

Example 3. The following judgments are classified into six categories taken to represent a continuum of opinion :—

Categories.

	I	II	III	IV	V	VI	Total
Judgment	8	21	42	51	17	5	144

Test the distribution versus normal distribution hypothesis.

Solution.

	Categories	I	II	III	IV	V	VI	Total
observed f_o	Judgments	8	21	42	51	17	5	144
expected f_e	Judgments	3	20	49	49	20	3	144

$(f_o - f_e) =$ 5 1 7 2 3 2

$(f_o - f_e)^2 =$ 25 1 49 4 9 4

$\frac{(f_o - f_e)^2}{f_e} =$ 8·33 ·05 1 ·082 ·45 1·33

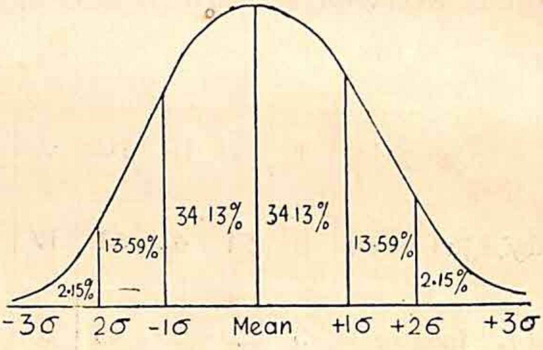
$$\chi^2 = 11{\cdot}242$$

degrees of freedom $= (6-1)\,(2-1) = 5 \times 1 = 5$

hence P is less than ·05.

In the first row of the above table the entries are the observed frequencies. To fill up the entries of the second row i. e. the expected frequencies (f_e) we consult table (9) of this part of this book. The base line of the normal curve (extended over 6σ) is to be divided in 6 segments each of 1σ. The simple problem is to calculate the number of frequencies that are expected to lie in each of these segments of normal distribution.

From the following figure we may be able to see the percentage of frequencies that are expected to lie in each segment of the normal distribution. From this we can calculate the expected frequencies in each segment if the total population be regarded as 144.



Curve 21

Segment	Percentage of frequencies	Expected frequencies if the total population is 144
Between 2σ and 3σ ...	2·15	...$\frac{2\cdot15}{100}\times 144=3$
Between 1σ and 2σ ...	13·59	...$\frac{13\cdot59}{100}\times 144=20$
Between Mean and 1σ...	34·13	...$\frac{34\cdot13}{100}\times 144=49$

Between Mean and -1σ ...	34·13	...$\frac{34\cdot13}{100}\times144=49$
Between -1σ and -2σ ...	13·59	...$\frac{13\cdot59}{100}+144=20$
Between -2σ and -3σ ...	2·15	...$\frac{2\cdot15}{100}\times144=3$
		Total $=144$

These expected frequencies of the third column are entered in the second row of the contingency table. When observed and expected frequencies are known we can calculate χ^2 ; *df* and P as shown above.

$$\chi^2=11\cdot242$$

$$df=5.$$

Assuming the confidence criteria ·05, we consult table of χ^2 at ·05 level.

The degrees of freedom are 5, the value of P at ·05 is 11·070. Our χ^2 is greater than this value and hence the divergence between the observed and expected results is significant. We can say with confidence that the deviation from the normal hypothesis is significant. Therefore we can reject to hypothesis of normal distribution.

The chi-square test when table entries are small

When table entries are large, estimates of probability given by χ^2-test are usually quite close to those obtained by direct methods. But when table entries are small (say five or less), and especially when the table is 2×2 fold (when the number of degrees of feedom is 1) the chi-square test is subject to considerable error. It is customary in such case to make a correction called the correction for continuity. Reasons for making this correction will be best understood from the example given below :

Example 4. Two sharp clicking sounds are presented in succession, the second being always more intense or less.

intense than the first. Presentation is in random order. In eight trials an observer is right six times. Is this result significant ?

Solution.

	Right	Wrong	Total
Observed (f_o)	6	2	8
Expected (f_e)	4	4	8

$$
\begin{array}{lll}
(f_o - f_e) = & 2 & 2 \\
\text{Correction} = -\cdot 5 & (1{\cdot}5) & (1{\cdot}5) \\
(f_o - f_e)^2 = & 2{\cdot}25 & 2{\cdot}25 \\
\dfrac{(f_o - f_e)^2}{f_e} = & \cdot 5625 & \cdot 5625 \\
\text{Chi}^2 \quad = & \cdot 5625 + & \cdot 5625
\end{array}
$$

$$= 1{\cdot}125$$

$$df = (2-1)\,(2-1) = 1$$

$P = \cdot 291$ (by interpolation in table of χ^2)

$$\frac{P}{2} = \cdot 145$$

In the above solution we introduced a correction of $-\cdot 5$ for each difference $(f_o - f_e)$. Rest procedure is same as in previous cases.

In applying the χ^2-test we assume that adjacent frequencies are connected by a continuous and smooth curve (like the normal curve) and are not discrete numbers. In 2×3 fold tables, however, in which the entries are small the curve is not continuous. Hence the deviation of 6 from 4 must be written as 1·5 i.e. $(5{\cdot}5 - 4)$ instead of 2i. e. $(6 - 4)$, since 5·5 is the lower limit of 6 in a continuous series. In like manner the deviation of 2 from 4 must be taken from the upper limit of 2, namely 2·5. Another change in procedure must be made in order to have the probability obtained from χ^2 agree with the direct determination of probability. P in the χ^2 table gives the probability of 6 or more right answers and 2 or less wrong answers that is it takes account of both ends of the probability curve. If we want only the probability of 6 or more right answers we should take $\frac{1}{2}$ of P, which as found above is ·145, in other words, if we repeat our test we should expect a score of 6 or better about 15 times in 100 trials. It is clear, therefore, that the obtained score is not significant.

When the expected entries in a 2×2 fold table are same as above, the formula for χ^2 may be written in a shorter form as below:

$$\chi^2 = 2\frac{(f_o - f_e)^2}{f_e}$$

The correction for continuity has little effect when table entries are large. But failure to use the correction even when numbers are fairly large may lead to some underestimation of the probability, hence it is generally wise to use it.

The χ^2-test when table entries are in percentages

The χ^2-test should not be used with percentage entries unless a correction for size of sample is made. This is of the fact because in dealing with probability the significance of an event depends upon its actual frequency and is not shown by its percentage occurrence. For a coin to fall heads 6 times in ten throws is not as significant as for the coin to fall head 60 times in 100 throws, although the percentage occurrence is same in both the cases.

Example 5. A coin was tossed 10 times before a subject who gave 7 correct judgments by reporting 'head' when the probability of a right judgment was $\frac{1}{2}$ in each instance. Calculate the χ^2.

Writing the entries in percentages, as below :—

Judgment	R	W	Total
(f_o)	70%	30%	100%
(f_e)	50%	50%	100%

	$(f_o-f_e)=20\%$	20%
Correction	(-5%) 15%	15%
	$(f_o-f_e)^2$ 225	225

$$\chi^2\% = \frac{2\,(225)}{50} = 9$$

$$\chi^2 = 9 \times \frac{10}{100} = \cdot 90$$

The unit here is 10%, so that 5% must be subtracted from each (f_o-f_e) difference.

Thus (70% $\sim$ 50%) is actually (65% $\sim$ 50%)
and (30% $\sim$ 50%) is (35% $\sim$ 50%).

It is clear that in order to bring χ^2 to its proper value in terms of original number we must multiply the 'percent' χ^2 by $\frac{10}{100}$ to give ·90. A χ^2 calculated from percentages must always be multiplied by $\frac{N}{100}$ (N = number of observations) in order to adjust it to the actual frequencies in the given sample.

We can easily test the significance of the hypothesis as in previous cases.

The χ^2-test of independence in contingency table

So far we have seen that χ^2 may be used in testing the agreement between the observed results and the expected results on some hypothesis. The χ^2 can also be used if we want to know the correspondence between attributes or traits which can be classified into two or more categories.

If we are dealing with peas (Pisum sativum) of different varities, they may be classified as to tall or dwarf, with green seeds or yellow seeds, with wrinkled seeds or round seeds and the correspondence in these attributes noted.

Example 6. The following table shows the result of inoculation against cholera. Is there any significant association between inoculation and attack ?

	Not attacked	Attacked	Total
Inoculated	431	5	436
Not inoculated	291	9	300
Total	722	14	736

Solution. In the following table 736 persons have been classified as to inoculated and attacked, inoculated and not attacked, not-inoculated and attacked and not-inoculated and not attacked. Reading down the first column we find that of 722 persons not atacked, 431 are inoculated and 291 not inoculated. Across the first row we find of 436 inoculated persons, 431 are not attacked and 5 attacked. The other columns and rows are interpreted in the same way. In order to compute χ^2 we must first calculate an 'independence value' for each cell in the contingency table.

'Independence values' are represented by figures in brackets within different cells. These figures in the brackets represent the expected frequencies in each cell. The method of calculating the independence value is shown in the following table. To illustrate with the first entry, there are 722 not attacked and 436 inoculated persons. If there were no association between 'not attacked' and 'inoculation' we should expect to find by chance $\frac{722 \times 436}{736}$ or 427·7 individuals in our group who are inoculated and not attacked. Similarly we can obtain the other expected frequencies. χ^2 can be calculated as shown below.

	Not-attacked	Attacked	Total
Inoculated	431 (427·7)	5 (8·3)	436
Not-inoculated	291 (294·3)	9 (5·7)	003
Total	732	14	736

Calculation of independence values

$$\frac{722 \times 436}{736} = 427{\cdot}1 \qquad \frac{14 \times 436}{736} = 8{\cdot}3$$

$$\frac{722 \times 300}{736} = 294{\cdot}3 \qquad \frac{14 \times 300}{736} = 5{\cdot}7$$

Calculation of χ^2.

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = \frac{(431 - 427{\cdot}7)^2}{427{\cdot}7} + \frac{(5 - 8{\cdot}3)^2}{8{\cdot}3}$$

$$+ \frac{(291 - 294{\cdot}3)^2}{294{\cdot}3} + \frac{(9 - 5{\cdot}7)^2}{5{\cdot}7}$$

$$=\frac{(3{\cdot}3)^2}{42{\cdot}77}+\frac{(3{\cdot}3)^2}{8{\cdot}3}+\frac{(3{\cdot}3)^2}{294{\cdot}3}+\frac{(3{\cdot}3)^2}{5{\cdot}7}=3{\cdot}27$$

$$\text{degrees of freedom}=df=(2-1)\ (2-1)=1$$

$$\therefore \quad P={\cdot}0706.$$

Thus if the hypothesis is true, our data, give a result which would be obtained about seven times in a hundred trials. This is infrequent, but not very infrequent. Moreover, the theoretical frequencies in the attacked column are not large. We should therefore be unjustified in rejecting the hypothesis on this evidence, but we can say that the data lend some colour to the supposition that the hypothesis is not correct.

To sum up, the χ^2 test shows that the data incline us, though not strongly, the belief that inoculation and attack are associated.

Example 7. An investigator into chocolate consumption divided the U. K. into eight areas and took a random sample from each, the individuals so obtained being classified as consumers or non-consumers of chocolate. His results were as follows :—

Area Number	1	2	3	4	5	6	7	8	Total
Cousumers	56 (55)	87 (81)	142 (152)	17 (69)	88 (90)	72 (72)	100 (95)	142 (144)	758
Non-Consumers	17 (18)	20 (26)	58 (48)	20 (22)	31 (29)	23 (23)	25 (30)	48 (46)	242
Total	73	107	200	91	119	95	125	190	1000

Do these results suggest that the consumption of chocolate varies from place to place ?

Let us take as our hypothesis H the supposition that it does not, i.e., that the two attributes in the above table are independent. The theoretical frequencies or expected frequencies f_e are then those shown in brackets and we have

$$\chi^2=\frac{1^2}{55}+\frac{6^2}{81}+14 \text{ similar terms}$$

$$=6{\cdot}28$$

The table has two rows and eight columns, and hence $df=(2-1)(8-1)=7$. From the table of χ^2, the point whose abscissa is 6·28 and ordinate 7 lies between the lines P=·7 and P=·5 very near the latter.

Thus there is no cause to suspect our hypothesis, and the data do not suggest that the consumption of chocolate varies from place to place, at least so far as this test is concerned.

SUMMARY

(1) $\chi^2=\Sigma\left[\frac{(f_o-f_e)^2}{f_e}\right]$

where f_o=frequency of occurrence of observed and experimentally determined facts.

f_e=Expected frequency of occurrence on some hypothesis.

(2) Short formula for χ^2 in 2×2 fold tables when expected frequancies are equal—

$$\chi^2=\frac{2(f_o-f_e)^2}{f_e}.$$

EXERCISES

1. 41 subjects are asked to express their attitude towards the proposition "should the United States join a Security Organization of Nations?" by marking F (favourable), I (indifferent) or U (unfavourable). Of the members in the group, twenty four marked F, twelve I, and twelve U. Do these results, indicate a significant trend of opinion ?

Ans. $\chi^2=6$: $df=2$; P=a bit less than ·05. The 'equal answer' hypothesis is rejected.

2. In answering a test of 100 true false items, a subject gets sixty right. From the chi-square test, determine whether this subject was merely guessing.

Hints. Since here we have a 2 × 2 fold table, we should use the correction for continuity (—·5)

Answer. $\chi^2=3{\cdot}62$

$df=1$

$P={\cdot}059$

$\frac{P}{2}={\cdot}0295$ or ·03 ; we may reject the null hypothesis with some confidence and conclude that our subject could not have been simply guessing.

3. Two investigators draw samples from the same town in order to estimate the number of persons falling, in the income groups 'Poorer', 'middle class' and 'well to do'. (The limits of the groups are defined in terms of money and are same for both investigators). Their results are as follows :—

Investigator	Income group			
	'Poorer'	'middle class'	'well to do'	Total
A	140	100	15	255
B	140	50	20	210
Totals	280	150	35	465

Show that the sampling technique of at least one of the investigators is suspect.

Ans. $\chi^2=13{\cdot}15$, $df=2$, $P=0{\cdot}0014$. This is rather low and we suspect the sampling to be non-random.

4. The following table represents the number of boys and the number of girls who chose each of the five possible answers to an item in the attitude scale.

	Strongly Approve	Approve	Indifferent	Disapprove	Strongly Disapprove	Total
Boys	25	30	10	25	10	100
Girls	10	15	5	15	15	60

Do these data indicate a significant sex difference in the attitude towards this question ? (Note : test the "independence (Null) hypothesis".

Ans, No, $\chi^2=7{\cdot}03$, $df=4$
and P is between ·20 and ·10.

5. 100 insurance agents have been classified into five groups—'excellent', 'very good', 'satisfactory', "poor", and 'very poor—by a company manager. Does this distribution of ratings differ significantly from that to be expected if agents ability is normally distributed ?

Ans. $\chi^2=172{\cdot}21$, $df=4$, P is much less than ·01. The hypothesis of a normal distribution of the ability to work as agents is rejected because the ratings differ significantly.

6. The table below shows the number of normals and abnormals who chose each of the three possible answers to an item on a neurotic questionnaire.

	Yes	No	?	Total
Normals	14	66	10	90
Abnormals	27	66	7	100
	41	132	17	190

Does the items differentiate between the two groups ? Test the independence hypothesis.

Ans. No, $\chi^2=4{\cdot}14$, $df=2$
and P is between ·20 and ·10.

TABLE OF SQUARES AND SQUARE ROOTS

Number	Square	Square Root	Number	Square	Square Root
1	1	1.000	51	26 01	7.141
2	4	1.414	52	27 04	7.211
3	9	1.732	53	28 09	7.280
4	16	2.000	54	29 16	7.348
5	25	2.236	55	30 25	7.416
6	36	2.449	56	31 36	7.483
7	49	2.646	57	32 49	7.550
8	64	2.828	58	33 64	7.616
9	81	3.000	59	34 81	7.681
10	1 00	3.162	60	36 00	7.746
11	1 21	3.317	61	37 21	7.810
12	1 44	3.464	62	38 44	7.874
13	1 69	3.606	63	39 69	7.937
14	1 96	3.742	64	40 96	8.000
15	2 25	3.873	65	42 25	8·062
16	2 56	4.000	66	43 56	8.124
17	2 89	4.123	67	44 89	8.185
18	3 24	4.243	68	46 24	8.246
19	3 61	4.359	69	47 61	8.307
20	4 00	4.472	70	49 00	8.367
21	4 41	4.583	71	50 41	8.426
22	4 84	4.690	72	51 84	8.485
23	5 29	4.796	73	53 29	8.544
24	5 76	4.899	74	54 76	8.602
25	6 25	5.000	75	56 25	8.660
26	6 76	5.099	76	57 76	8.718
27	7 29	5.196	77	59 29	8.775
28	7 84	5.292	78	60 84	8.832
29	8 41	5.385	79	62 41	8.888
30	9 00	5.477	80	64 00	8.944
31	9 61	5.568	81	65 61	9.000
32	10 24	5.657	82	67 24	9.055
33	10 89	5.745	83	68 89	9.110
34	11 56	5.831	84	70 56	9.165
35	12 25	5.916	85	72 25	9.220
36	12 96	6.000	86	73 96	9.274
37	13 69	6.083	87	75 69	9.327
38	14 44	6.164	88	77 44	9.381
39	15 21	6.245	89	79 21	9.434
40	16 00	6.325	90	81 00	9.487
41	16 81	6.403	91	82 81	9.539
42	17 64	6.481	92	84 64	9.592
43	18 49	6.557	93	86 49	9.644
44	19 36	6.633	94	88 36	9.695
45	20 25	6.708	95	90 25	9.747
46	21 16	6.782	96	92 16	9.798
47	22 09	6.856	97	94 09	9.849
48	23 04	6.928	98	96 04	9·899
49	24 01	7.000	99	98 01	9·950
50	25 00	7.071	100	1 00 00	10·000

Number	Square	Square Root	Number	Square	Square Root
101	1 02 01	10.050	151	2 28 01	12.288
102	1 04 04	10.100	152	2 31 04	12.329
103	1 06 09	10.149	153	2 34 09	12.369
104	1 08 16	10.198	154	2 37 16	12.410
105	1 10 25	10.247	155	2 40 25	12.450
106	1 12 36	10.296	156	2 43 36	12.490
107	1 14 49	10.344	157	2 46 49	12.530
108	1 16 64	10.392	158	2 49 64	12.570
109	1 18 81	10.440	159	2 52 81	12.610
110	1 21 00	10.488	160	2 56 00	12.649
111	1 23 21	10.536	161	2 59 21	12.689
112	1 25 44	10.583	162	2 62 44	12.728
113	1 27 69	10.630	163	2 65 69	12.767
114	1 29 96	10.677	164	2 68 96	12.806
115	1 32 25	10.724	165	2 72 25	12.845
116	1 34 56	10.770	166	2 75 56	12.884
117	1 36 89	10.817	167	2 78 89	12.923
118	1 39 24	10.863	168	2 82 24	12.961
119	1 41 61	10.909	169	2 85 61	13.000
120	1 44 00	10.954	170	2 89 00	13.038
121	1 46 41	11.000	171	2 92 41	13.077
122	1 48 84	11.045	172	2 95 84	13.115
123	1 51 29	11.091	173	2 99 29	13.153
124	1 53 76	11.136	174	3 02 76	13.191
125	1 56 25	11.180	175	3 06 25	13.229
126	1 58 76	11.225	176	3 09 76	13.266
127	1 61 29	11.269	177	3 13 29	13.304
128	1 63 84	11.314	178	3 16 84	13.342
129	1 66 41	11.358	179	3 20 41	13.379
130	1 69 00	11.402	180	3 24 00	13.416
131	1 71 61	11.446	181	3 27 61	13.454
132	1 74 24	11.489	182	3 31 24	13.491
133	1 76 89	11.533	183	3 34 89	13.528
134	1 79 56	11.576	184	3 38 56	13.565
135	1 82 25	11.619	185	3 42 25	13.601
136	1 84 96	11.662	186	3 45 96	13.638
137	1 87 69	11.705	187	3 49 69	13.675
138	1 90 44	11.747	188	3 53 44	13.711
139	1 93 21	11.790	189	3 57 21	13.748
140	1 96 00	11.832	190	3 61 00	13.784
141	1 98 81	11.874	191	3 64 81	13.820
142	2 01 64	11.916	192	3 68 64	13.856
143	2 04 49	11.958	193	3 72 49	13.892
144	2 07 36	12.000	194	3 76 36	13.928
145	2 10 25	12.042	195	3 80 25	13.964
146	2 13 16	12.083	196	3 84 16	14.000
147	2 16 09	12.124	197	3 88 09	14.036
148	2 19 04	12.166	198	3 92 04	14.071
149	2 22 01	12.207	199	3 96 01	14.107
150	2 25 00	12.247	200	4 00 00	14.142

Number	Square	Square Root	Number	Square	Square Root
201	4 04 01	14.177	251	6 30 01	15.843
202	4 08 04	14.213	252	6 35 04	15.875
203	4 12 09	14.248	253	6 40 09	15.906
204	4 16 16	14.283	254	6 45 16	15.937
205	4 20 25	14.318	255	6 50 25	15.969
206	4 24 36	14.353	256	6 55 36	16.000
207	4 28 49	14.387	257	6 60 49	16.031
208	4 32 64	14.422	258	6 65 64	16.062
209	4 36 81	14.457	259	6 70 81	16.093
210	4 41 00	14.491	260	6 76 00	16.125
211	4 45 21	14.526	261	6 81 21	16.155
212	4 49 44	14.560	262	6 86 44	16.186
213	4 53 69	14.595	263	6 91 69	16.217
214	4 57 96	14.629	264	6 96 96	16.248
215	4 62 25	14.663	265	7 02 25	16.279
216	4 66 56	14.697	266	7 07 56	16.310
217	4 70 89	14.731	267	7·12 89	16.340
218	4 75 24	14.765	268	7 18 24	16.371
219	4 79 61	14.799	269	7 23 61	16.401
220	4 84 00	14.832	270	7 29 00	16.432
221	4 88 41	14.866	271	7 34 41	16.462
222	4 92 84	14.900	272	7 39 84	16.492
223	4 97 29	14.933	273	7 45 29	16.523
224	5 01 76	14.967	274	7 50 76	16.553
225	5 06 25	15.000	275	7 56 25	16.583
226	5 10 76	15.033	276	7 61 76	16.613
227	5 15 29	15.067	277	7 67 29	16.643
228	5 19 84	15.100	278	7 72 84	16.673
229	5 24 41	15.133	279	7 78 41	16.703
230	5 29 00	15.166	280	7 84 00	16.733
231	5 33 61	15.199	281	7 89 61	16.763
232	5 38 24	15.232	282	7 95 24	16.793
233	5 42 89	15.264	283	8 00 89	16.823
234	5 47 56	15.297	284	8 06 56	16.852
235	5 52 25	15.330	285	8 12 25	16.882
236	5 56 96	15.362	286	8 17 96	16.912
237	5 61 69	15.395	287	8 23 69	16.941
238	5 66 44	15.427	288	8 29 44	16.971
239	5 71 21	15.460	289	8 35 21	17.000
240	5 76 00	15.492	290	8 41 00	17.029
241	5 80 81	15.524	291	8 46 81	17.059
242	5 85 64	15.556	292	8 52 64	17.088
243	5 90 49	15.588	293	8 58 49	17·117
244	5 95 36	15.620	294	8 64 36	17.146
245	6 00 25	15.652	295	8 70 25	17.176
246	6 05 16	15.684	296	8 76 16	17.205
247	6 10 09	15.716	297	8 82 09	17.234
248	6 15 04	15.748	298	8 88 04	17.263
249	6 20 01	15.780	299	8 94 01	17.292
250	6 25 00	15.811	300	9 00 00	17.321

Number	Square	Square Root	Number	Square	Square Root
301	9 06 01	17.349	351	12 32 01	18.735
302	9 12 04	17.378	352	12 39 04	18.762
303	9 18 09	17.407	353	12 46 09	18.788
304	9 24 16	17.436	354	12 53 16	18.815
305	9 30 25	17.464	355	12 60 25	18.841
306	9 36 36	17.493	356	12 67 36	18.868
307	9 42 49	17.521	357	12 74 49	18.894
308	9 48 64	17.550	358	12 81 64	18.921
309	9 54 81	17.578	359	12 88 81	18.947
310	9 61 00	17.607	360	12 96 00	18.974
311	9 67 21	17.635	361	13 03 21	19.000
312	9 73 44	17.664	362	13 10 44	19.026
313	9 79 69	17.692	363	13 17 69	19.053
314	9 85 96	17.720	364	13 24 96	19.079
315	9 92 25	17.748	365	13 32 25	19.105
316	9 98 56	17.776	366	13 39 56	19.131
317	10 04 89	17.804	367	13 46 89	19.157
318	10 11 24	17.833	368	13 54 24	19.183
319	10 17 61	17.861	369	13 61 61	19.209
320	10 24 00	17.889	370	13 69 00	19.235
321	10 30 41	17.916	371	13 76 41	19.261
322	10 36 84	17.944	372	13 83 84	19.287
323	10 43 29	17.972	373	13 91 29	19.313
324	10 49 76	18.000	374	13 98 76	19.339
325	10 56 25	18.028	375	14 06 25	19.363
326	10 62 76	18.055	376	14 13 76	19.391
327	10 69 29	18.083	377	14 21 29	19.416
328	10 75 84	18.111	378	14 28 84	19.442
329	10 82 41	18.138	379	14 36 41	19.468
330	10 89 00	18.166	380	14 44 00	19.494
331	10 95 61	18.193	381	14 51 61	19.519
332	11 02 24	18.221	382	14 59 24	19.545
333	11 08 89	18.248	383	14 66 89	19.570
334	11 15 56	18.276	384	14 74 56	19.596
335	11 22 25	18.303	385	14 82 25	19.621
336	11 28 96	18.330	386	14 89 96	19.647
337	11 35 69	18,358	387	14 97 69	19.672
338	11 42 44	18.385	388	15 05 44	19.698
339	11 49 21	18.412	389	15 13 21	19.723
340	11 56 00	18.439	390	15 21 00	19.748
341	11 62 81	18.466	391	15 28 81	19.774
342	11 69 64	18.493	392	15 36 64	19.799
343	11 76 49	18.520	393	15 44 49	19.824
344	11 83 36	18.547	394	15 52 36	19.849
345	11 90 25	18.574	395	15 60 25	19.875
346	11 97 16	18.601	396	15 68 16	19.900
347	12 04 09	18.628	397	15 76 09	19,925
348	12 11 04	18,655	398	15 84 04	19.950
349	12 18 01	18.682	399	15 92 01	19.975
350	12 25 00	18.708	400	16 00 00	20.000

Number	Square	Square Root	Number	Square	Square Root
401	16 08 01	20.025	451	20 34 01	21.237
402	16 16 04	20.050	452	20 43 04	21.260
403	16 24 09	20.075	453	20 52 09	21.284
404	16 32 16	20.100	454	20 61 16	21.307
405	16 40 25	20.125	455	20 70 25	21.331
406	16 48 36	20.149	456	20 79 36	21.354
407	16 56 49	20.174	457	20 88 49	21.378
408	16 64 64	20.199	458	20 97 64	21.401
409	16 72 81	20.224	459	21 06 81	21.424
410	16 81 00	20.248	460	21 16 00	21.448
411	16 89 21	20.273	461	21 25 21	21.471
412	16 97 44	20.298	462	21 34 44	21.494
413	17 05 69	20.322	463	21 43 69	21.517
414	17 13 96	20.347	464	21 52 96	21.541
415	17 22 25	20.372	465	21 62 25	21.564
416	17 30 56	20.396	466	21 71 56	21.587
417	17 38 89	20.421	467	21 80 89	21.610
418	17 47 24	20.445	468	21 90 24	21.633
419	17 55 61	20.469	469	21 99 61	21.656
420	17 64 00	20.494	470	22 09 00	21.679
421	17 72 41	20.518	471	22 18 41	21.703
422	17 80 84	20.543	472	22 27 84	21.726
423	17 89 29	20.567	473	22 37 29	21.749
424	17 97 76	20.591	474	22 46 76	21.772
425	18 06 25	20.616	475	22 56 25	21.794
426	18 14 76	20.640	476	22 65 76	21.817
427	18 23 29	20.664	477	22 75 29	21.840
428	18 31 84	20.688	478	22 84 84	21.863
429	18 40 41	20.712	479	22 94 41	21.886
430	18 49 00	20.736	480	23 04 00	21.909
431	18 57 61	20.761	481	23 13 61	21.932
432	18 66 24	20.785	482	23 23 24	21.954
433	18 74 89	20.809	483	23 32 89	21.977
434	18 83 56	20.833	484	23 42 56	22.000
435	18 92 25	20.857	485	23 52 25	22.023
436	19 00 96	20.881	486	23 61 96	22.045
437	19 09 69	20.905	487	23 71 69	22.068
438	19 18 44	20.928	488	23 81 44	22.091
439	19 27 21	20.952	489	23 91 21	22.113
440	19 36 00	20.976	490	24 01 00	22.136
441	19 44 81	21.000	491	24 10 81	22.159
442	19 53 64	21.024	492	24 20 64	22.181
443	19 62 49	21.048	493	24 30 49	22.204
444	19 71 36	21.071	494	24 40 36	22.226
445	19 80 25	21.095	495	24 50 25	22.249
446	19 89 16	21.119	496	24 60 16	22.271
447	19 98 09	21.142	497	24 70 09	22.293
448	20 07 04	21.166	498	24 80 04	22.316
449	20 16 01	21.190	499	24 90 01	22.338
450	20 25 00	21.213	500	25 00 00	22.361

Number	Square	Square Root	Number	Square	Square Root
501	25 10 01	22.383	551	30 36 01	23.473
502	25 20 04	22.405	552	30 47 04	23.495
503	25 30 09	22.428	553	30 58 09	23.516
504	25 40 16	22.450	554	30 69 16	23.537
505	25 50 25	22.472	555	30 80 25	23.558
506	25 60 36	22.494	556	30 91 36	23.580
507	25 70 49	22.517	557	31 02 49	23.601
508	25 80 64	22.539	558	31 13 64	23.622
509	25 90 81	22.561	559	31 24 81	23.643
510	26 01 00	22.583	560	31 36 00	23.664
511	26 11 21	22.605	561	31 47 21	23.685
512	26 21 44	22.627	562	31 58 44	23.707
513	26 31 69	22.650	563	31 69 69	23.728
514	26 41 96	22.672	564	31 80 96	23.749
515	26 52 25	22.694	565	31 92 25	23.770
516	26 62 56	22.716	566	32 03 56	23.791
517	26 72 89	22.738	567	32 14 89	23.812
518	26 83 24	22.760	568	32 26 24	23.833
519	26 93 61	22.782	569	32 37 61	23.854
520	27 04 00	22.804	570	32 49 00	23.875
521	27 14 41	22.825	571	32 60 41	23.896
522	27 24 84	22.847	572	32 71 84	23.917
523	27 35 29	22.869	573	32 83 29	23.937
524	27 45 76	22.891	574	32 94 76	23.958
525	27 56 25	22.913	575	33 06 25	23.979
526	27 66 76	22.935	576	33 17 76	24.000
527	27 77 29	22.956	577	33 29 29	24.021
528	27 87 84	22.978	578	33 40 84	24.042
529	27 98 41	23.000	579	33 52 41	24.062
530	28 09 00	23.022	580	33 64 00	24.083
531	28 19 61	23.043	581	33 75 61	24.104
532	28 30 24	23.065	582	33 87 24	24.125
533	28 40 89	23.087	583	33 98 89	24.145
534	28 51 56	23.108	584	34 10 56	24.166
535	28 62 25	23.130	585	34 22 25	24.187
536	28 72 96	23.152	586	34 33 96	24.207
537	28 83 69	23.173	587	34 45 69	24.228
538	28 94 44	23.195	588	34 57 44	24.249
539	29 05 21	23.216	589	34 69 21	24.269
540	29 16 00	23.238	590	34 81 00	24.290
541	29 26 81	23.259	591	34 92 81	24.310
542	29 37 64	23.281	592	35 04 64	24.331
543	29 48 49	23.302	593	35 16 49	24.352
544	29 59 36	23.324	594	35 28 36	24.372
545	29 70 25	23.345	595	35 40 25	24.393
546	29 81 16	23.367	596	35 52 16	24.413
547	29 92 09	23.388	597	35 64 09	24.434
548	30 03 04	23.409	598	35 76 04	24.454
549	30 14 01	23.431	599	35 88 01	24.474
550	30 25 00	23.452	600	36 00 00	24.495

Number	Square	Square Root	Number	Square	Square Root
601	36 12 01	24.515	651	42 38 01	25.515
602	36 24 04	24.536	652	42 51 04	25.534
603	36 36 09	24.556	653	42 64 09	25.554
604	36 48 16	24.576	654	42 77 16	25.573
605	36 60 25	24.597	655	42 90 25	25.593
606	36 72 36	24.617	656	43 03 36	25.612
607	36 84 49	24.637	657	43 16 49	25.632
608	36 96 64	24.658	658	43 29 64	25.652
609	37 08 81	24.678	659	43 42 81	25.671
610	37 21 00	24.698	660	43 56 00	25.690
611	37 33 21	24.718	661	43 69 21	25.710
612	37 45 44	24.739	662	43 82 44	25.729
613	37 57 69	24.759	663	43 95 69	25.749
614	37 69 96	24.779	664	44 08 96	25.768
615	37 82 25	24.799	665	44 22 25	25.788
616	37 94 56	24.819	666	44 35 56	25.807
617	38 06 89	24.839	667	44 48 89	25.826
618	38 19 24	24.860	668	44 62 24	25.846
619	38 31 61	24.880	669	44 75 61	25.865
620	38 44 00	24.900	670	44 89 00	25.884
621	38 56 41	24.920	671	45 02 41	25.904
622	38 68 84	24.940	672	45 15 84	25.923
623	38 81 29	24.960	673	45 29 29	25.942
624	38 93 76	24.980	674	45 42 76	25.962
625	39 06 25	25.000	675	45 56 25	25.981
626	39 18 76	25.020	676	45 69 76	26.000
627	39 31 29	25.040	677	45 83 29	26.019
628	39 43 84	25.060	678	45 96 84	26.038
629	39 56 41	25.080	679	46 10 41	26.058
630	39 69 00	25.100	680	46 24 00	26.077
631	39 81 61	25.120	681	46 37 61	26.096
632	39 94 24	25.140	682	46 51 24	26.115
633	40 06 89	25.159	683	46 64 89	26.134
634	40 19 56	25.179	684	46 78 56	26.153
635	40 32 25	25.199	685	46 92 25	26.173
636	40 44 96	25.219	686	47 05 96	26.192
637	40 57 69	25.239	687	47 19 69	26.211
638	40 70 44	25.259	688	47 33 44	26.230
639	40 83 21	25.278	689	47 47 21	26.249
640	40 96 00	25.298	690	47 61 00	26.268
641	41 08 81	25.318	691	47 74 81	26.287
642	41 21 64	25.338	692	47 88 64	26.306
643	41 34 49	25.357	693	48 02 49	26.325
644	41 47 36	25.377	694	48 16 36	26.344
645	41 60 25	25.397	695	48 30 25	26.363
646	41 73 16	25.417	696	48 44 16	26.382
647	41 86 09	25.436	697	48 58 09	26.401
648	41 99 04	25.456	698	48 72 04	26.420
649	42 12 01	25.475	699	48 86 01	26.439
650	42 25 00	25.495	700	49 00 00	26.458

Number	Square	Square Root	Number	Square	Square Root
701	49 14 01	26.476	751	56 40 01	27.404
702	49 28 04	26.495	752	56 55 04	27.423
703	49 42 09	26.514	753	56 70 09	27.441
704	49 56 16	26.533	754	56 85 16	27.459
705	49 70 25	26.552	755	57 00 25	27.477
706	49 84 36	26.571	756	57 15 36	27.495
707	49 98 49	26.589	757	57 30 49	27.514
708	50 12 64	26.608	758	57 45 64	27.532
709	50 26 81	26.627	759	57 60 81	27.550
710	50 41 00	26.646	760	57 76 00	27.568
711	50 55 21	26.665	761	57 91 21	27.586
712	50 69 44	26.683	762	58 06 44	27.604
713	50 83 69	26.702	763	58 21 69	27.622
714	50 97 96	26.721	764	58 36 96	27.641
715	51 12 25	26.739	765	58 52 25	27.659
716	51 26 56	26.758	766	58 67 56	27.677
717	51 40 89	26.777	767	58 82 89	27.695
718	51 55 24	26.796	768	58 98 24	27.713
719	51 69 61	26.814	769	59 13 61	27.731
720	51 84 00	26.833	770	59 29 00	27.749
721	51 98 41	26.851	771	59 44 41	27.767
722	52 12 84	26.870	772	59 59 84	27.785
723	52 27 29	26.889	773	59 75 29	27.803
724	52 41 76	26.907	774	59 90 76	27.821
725	52 56 25	26.926	775	60 06 25	27.839
726	52 70 76	26.944	776	60 21 76	27.857
727	52 85 29	26.963	777	60 37 29	27.875
728	52 99 84	26.981	778	60 52 84	27.893
729	53 14 41	27.000	779	60 68 41	27.911
730	53 29 00	27.019	780	60 84 00	27.928
731	53 43 61	27.037	781	60 99 61	27.946
732	53 58 24	27.055	782	61 15 24	27.964
733	53 72 89	27.074	783	61 30 89	27.982
734	53 87 56	27.092	784	61 46 56	28.000
735	54 02 25	27.111	785	61 62 25	28.018
736	54 16 96	27.129	786	61 77 96	28.036
737	54 31 69	27.148	787	61 93 69	28.054
738	54 46 44	27.166	788	62 09 44	28.071
739	54 61 21	27.185	789	62 25 21	28.089
740	54 76 00	27.203	790	62 41 00	28.107
741	54 90 81	27.221	791	62 56 81	28.125
742	55 05 64	27.240	792	62 72 64	28.142
743	55 20 49	27.258	793	62 88 49	28.160
744	55 35 36	27.276	794	63 04 36	28.178
745	55 50 25	27.295	795	63 20 25	28.196
746	55 65 16	27.313	796	63 36 16	28.213
747	55 80 09	27.331	797	63 52 09	28.231
748	55 95 04	27.350	798	63 68 04	28.249
749	56 10 01	27.368	799	63 84 01	28.267
750	56 25 00	27.386	800	64 00 00	28.284

Number	Square	Square Root	Number	Square	Square Root
801	64 16 01	28.302	851	72 42 01	29.172
802	64 32 04	28.320	852	72 59 04	29.189
803	64 48 09	28.337	853	72 76 09	29.206
804	64 64 16	28.355	854	72 93 16	29.223
805	64 80 25	28.373	855	73 10 25	29.240
806	64 96 36	28.390	856	73 27 36	29.257
807	65 12 49	28.408	857	73 44 49	29.275
808	65 28 64	28.425	858	73 61 64	29.292
809	65 44 81	28.443	859	73 78 81	29.309
810	65 61 00	28.460	860	73 96 00	29.326
811	65 77 21	28.478	861	74 13 21	29.343
812	65 93 44	28.496	862	74 30 44	29.360
813	66 09 69	28.513	863	74 47 69	29.377
814	66 25 96	28.531	864	74 64 96	29.394
815	66 42 25	28.548	865	74 82 25	29.411
816	66 58 56	28.566	866	74 99 56	29.428
817	66 74 89	28.583	867	75 16 89	29.445
818	66 91 24	28.601	868	75 34 24	29.462
819	67 07 61	28.618	869	75 51 61	29.479
820	67 24 00	28.636	870	75 69 00	29.496
821	67 40 41	28.653	871	75 86 41	29.513
822	67 56 84	28.671	872	76 03 84	29.530
823	67 73 29	28.688	873	76 21 29	29.547
824	67 89 76	28.705	874	76 38 76	29.563
825	68 06 25	28.723	875	76 56 25	29.580
826	68 22 76	28.740	876	76 73 76	29.597
827	68 39 29	28.758	877	76 91 29	29.614
828	68 55 84	28.775	878	77 08 84	29.631
829	68 72 41	28.792	879	77 26 41	29.648
830	68 89 00	28.810	880	77 44 00	29.665
831	69 05 61	28.827	881	77 61 61	29.682
832	69 22 24	28.844	882	77 79 24	29.698
833	69 38 89	28.862	883	77 96 89	29.715
834	69 55 56	28.879	884	78 14 56	29.732
835	69 72 25	28.896	885	78 32 25	29.749
836	69 88 96	28.914	886	78 49 96	29.766
837	70 05 69	28.931	887	78 67 69	29.783
838	70 22 44	28.948	888	78 85 44	29.799
839	70 39 21	28.965	889	79 03 21	29.816
840	70 56 00	28.983	890	79 21 00	29.833
841	70 72 81	29.000	891	79 38 81	29.850
842	70 89 64	29.017	892	79 56 64	29.866
843	71 06 49	29.034	893	79 74 49	29.883
844	71 23 36	29.052	894	79 92 36	29.900
845	71 40 25	29.069	895	80 10 25	29.916
846	71 57 16	29.086	896	80 28 16	29.933
847	71 74 09	29.103	897	80 46 09	29.950
848	71 91 04	29.120	898	80 64 04	29.967
849	72 08 01	29.138	899	80 82 01	29.983
850	72 25 00	29.155	900	81 00 00	30.000

Number	Square	Square Root
901	81 18 01	30.017
902	81 36 04	30.033
903	81 54 09	30.050
904	81 72 16	30.067
905	81 90 25	30.083
906	82 08 36	30.100
907	82 26 49	30.116
908	82 44 64	30.133
909	82 62 81	30.150
910	82 81 00	30.166
911	82 99 21	30.183
912	83 17 44	30.199
913	83 35 69	30.216
914	83 53 96	30.232
915	83 72 25	30.249
916	83 90 56	30.265
917	84 08 89	30.282
918	84 27 24	30.299
919	84 45 61	30.315
920	84 64 00	30.332
921	84 82 41	30.348
922	85 00 84	30.364
923	85 19 29	30.381
924	85 37 76	30.397
925	85 56 25	30.414
926	85 74 76	30.430
927	85 93 29	30.447
928	86 11 84	30.463
929	86 30 41	30.480
930	86 49 00	30.496
931	86 67 61	30.512
932	86 86 24	30.529
933	87 04 89	30.545
934	87 23 56	30.561
935	87 42 25	30.578
936	87 60 96	30.594
937	87 79 69	30.610
938	87 98 44	30.627
939	88 17 21	30.643
940	88 36 00	30.659
941	88 54 81	30.676
942	88 73 64	30.692
943	88 92 49	30.708
944	89 11 36	30.725
945	89 30 25	30.741
946	89 49 16	30.757
947	89 68 09	30.773
948	89 87 04	30.790
949	90 06 01	30.806
950	90 25 00	30.822
951	90 44 01	30.838
952	90 63 04	30.854
953	90 82 09	30.871
954	91 01 16	30.887
955	91 20 25	30.903
956	91 39 36	30.919
957	91 58 49	30.935
958	91 77 64	30.952
959	91 96 81	30.968
960	92 16 00	30.984
961	92 35 21	31.000
962	92 54 44	31.016
963	92 73 69	31.032
964	92 92 96	31.048
965	93 12 25	31.064
966	93 31 56	31.081
967	93 50 89	31.097
968	93 70 24	31.113
969	93 89 61	31.129
970	94 09 00	31.145
971	94 28 41	31.161
972	94 47 84	31.177
973	94 67 29	31.193
974	94 86 76	31.209
975	95 06 25	31.225
976	95 25 76	31.241
977	95 45 29	31.257
978	95 64 84	31.273
979	95 84 41	31.289
980	96 04 00	31.305
981	96 23 61	31.321
982	96 43 24	31.337
983	96 62 89	31.353
984	96 82 56	31.369
985	97 02 25	31.385
986	97 21 96	31.401
987	97 41 69	31.417
988	97 61 44	31.432
989	97 81 21	31.448
990	98 01 00	31.464
991	98 20 81	31.480
992	98 40 64	31.496
993	98 60 49	31.512
994	98 80 36	31.528
995	99 00 25	31.544
996	99 20 16	31.559
997	99 40 09	31.575
998	99 60 04	31.591
999	99 80 01	31.607
1000	100 00 00	31.623

BIBLIOGRAPHY

(A) Books on Measurement:

(1) ANASTASI, A. : 'Psychological Testing'; *M/s Macmillan Co., New York, 1954.*

(2) ALLEN, W. C. : 'Cumulative Pupil Records'; *Bureau of Publications, Teachers College, Columbia University, New York (1943).*

(3) BASU., A. N. : 'School Record'; *Ministry of Education, Government of India, 1955.*

(4) BALLARD, P. B. : 'Mental Tests'; *University of London Press, London.*

(5) BALLARD, P. B. : 'Group Tests of Intelligence'; *University of London Press, London.*

(6) BALLARD, P. B. : 'The New Examiner': *University of London Press, London, 1923.*

(7) BHATIA, C. M. : 'Performance test of Intelligence under Indian Conditions'; *Geoffrey Cumberlege, Oxford University Press, 1954.*

(8) BLOOM, S. B. : 'Taxonomy of Educational Objectives'; *Longmans, Green & Co., New York, London, Toronto, 1956.*

(9) BURT, C. L. : 'Age, Ability and Aptitude'; *University London Inst. Educ. Stud. Educ., No. 6 'The Problems of Secondary Education To-day'; London, Evans, 1954.*

(10) BURT, C. : 'Mental & Scholastic Tests'; *Staples Press Ltd., Cavendish Place, London; W. I., 1921.*

(11) CATTELL, R. B. : 'A Guide to Mental Testing'; *University of London Press, London, 1936.*

(12) EARLE, F. M. : 'Reconstruction in the Secondary School'; *University of London Press, London, 1943.*

(13) GREEN, H. A. : 'Measurement and Evaluation in Secondary School'; *Longmans, Green & Co., New York, 1955.*

(14) GERBERICH, I. R. : 'Specimen Objective Test Items'; *Longmans, Green & Co. New York, London, Toronto, 1956.*

(15) GOODENOUGH, F. L : 'Mental Testing'; *Rinehart & Co., Inc., Publishers, New York, 1949.*

(16) GULLIKSEN, H. : 'Theory of Mental Tests'; *John Wiley & Sons, Inc., New York.*

(17) GUPTA, R. K. : 'An attempt to construct and standardise Test in General Science with reference to some of the aims of teaching the subject at ordinary level, G. C. E.' *M. A. Thesis, London University.*

(18) HAWKES, H. E. ; LINDQUIST, E. F. AND MANN, C. R. : 'The Construction and Use of Achievement Examinations' ;
Boston Houghton Miffin Co., 1963.

(19) HUNTER, G. W. AND KNAPP. A. : 'Hunter and Knapp, Mastery Test in General Science' ;
American Book Co., New York, 1934.

(20) JORDON, A. M. : 'Measurement in Education' ;
McGraw Hill Book Co., Inc., New York, Toronto, London, 1953.

(21) KNIGHT, R. : 'Intelligence Test' ;
Methuen, London, 1933.

(22) KUDER, G. F., AND RICHARDSON, M. W. : 'The Theory of Estimation of Test Reliability' ;
'Psychometricka', 1937.

(23) KELLY, T. L. : 'The Reliability of Test Scores' ;
Journal of Educational Research, 1921, 3, 327.

(24) LINDQUIST, E. F. : 'Educational Measurement' ;
American Council on Education, Washington, D. C.

(25) LESTER SMITH, W. O. : 'Education' ;
Penguin Books, 1957.

(26) MICHEELS, W. J., KARNES, M. R. : 'Measuring Educational Achievement' ;
McGraw Hill Book Co., Inc., New York, Toronto, London, 1950.

(27) MENZEL. E. W. : 'Suggestions for the use of New-Type Test in India' ;
Geoffrey Cumberlege, Oxford University Press.

(28) MEALINGS R. J. : 'General Science Test' (unpublished) ;
Birmingham University.

(29) MURSELL, J. L. : 'Psychological Testing' ;
Longman Green & Co., London, 1950.

(30) PEEL, E. A. AND RUTTER, D. : 'The Predictive Value of the Entrance Examination as judged by the School Certificate Examination' ;
Brit, J. Educ. Psychology, 21, 30–35, 1951.

(31) PEEL, E. A. : 'Assessment of Interest in Practical Topics' ;
Brit, J. Educ. Psychology, 1948, 18, 41–7.

(32) PIAGET, J. : The Psychology of Intelligence' ;
Routledge & Kegan Paul Ltd., London, 1956.

(33) REMMERS, H. H. AND GAGE, N. L. : 'Educational Measurement and Evaluation' ;
Harper and Brothers, Publishers, New York, 1954.

(34) 'Educational Tests and Their Uses' ;
'Review of Educational Research', 5 : 482, 1935.

(35) RAWAT, D. S. :'Vidyalaya Mai Mapan Awar Moolyankan' ;
Gaya Pd. & Sons, Agra, 1962.

(36) RAWAT, D. S. : 'A Comparative study of the Standard of General Science, in certain Secondary Schools in England, and in U. P. (India) for pupils of a chosen Age Range' ; M. A. (Education Thesis) ;
Birmingham University, 1958.

(37) ROSS. C. C. : 'Measurement in To-days' Schools' ;
New York, Prentice Hall, Inc., 1955.

(38) SCHONELL, F. J. AND SCHONELL. E. F. : 'Diagnostic and Attainment Testing' ;
Oliver & Boyd, Edinburgh, Tweeddale Court, London, 39 A Welback Street, W. I.

(39) SPEARMAN, C. & JONES LLWYNN : 'Human Ability' ;
London, Macmillan & Co., Ltd., 1951.

(40) THORNDIKF, R. L. & HAGEN, E. : 'Measurement & Education in Psychology and Education' ;
John Wiley & Sons Ltd., London, 1961.

(41) TERMAN, L. M. & MERRILL, M. A. : 'Measuring Intelligence';
Harrap, London, 1937.

(42) VERNON, P. E. : 'The Structure of Human Abilities' ;
London, Mathuen & Co. Ltd., New York, John Wiley and Sons Inc., 1950.

(43) VERNON, P. E. : 'Secondary School Selection' ;
Methuen & Co. Ltd., 36 Essex Street, W. C. 2, 1957.

(44) VALENTINE, C. W. AND EMMETT, W. G. : 'The Reliability of Examinations' ;
London, University of London Press, 1932.

(45) WHIPPLE, G. M. : 'Intelligence, Its Nature & Nurture' ; The thirtyninth Year book of National Society for the study of Education, Part I & II.
Public School Publishing Co., Bloomington, Illinois.

(B) Books on Statistics :

(1) BROOM, M. E. : 'Educational Statistics' ;
New York, Cincinnati, Chicago, American Book Co., Boston, Atlanta, Dallas, San Francisco.

(2) EDWARDS, A. L. : 'Experimental Design in Psychological Research' ;
New York, Rinehart & Co., 1950.

(3) GARRETT, HENREY, E. : 'Statistics in Psychology and Education' ;
Longmans Green & Co., New York, London, Toronto.

(4) GUILFORD, J. P. : 'Fundamental Statistics in Psychology and Education' ; (3rd Ed.)
New York, McGraw Hill Book Co., Inc., 1956.

(5) GUILFORD, J. P. : 'Psychometric Methods' ;
McGraw Hill, New York, 1954.

(6) LINDQUIST, E. F. : 'Statistical Analysis in Educational Research' ; *Harrap, 1940.*

(7) MCNEMAR, Q. : 'Psychological Statistics' ;
John Wiley & Sons. Inc., New York ; Chapman & Hall Ltd London, 1955.

(8) KUDER, G. F. AND RICHARDSON, M. W. : 'Psychometrika' ; (The Theory of Estimation of Test Reliability), 1937, 2, 151–160.

(9) RICHARDSON, M. W. AND KUDER, G. F. : 'Journal of Educational Psychology' ; (The Calculation of Test Reliability Coefficient Based upon the methods of Rational Equivalence), 1939, 30, 681-687.

(10) RAWAT, D. S. : 'Statistics in Education' ;
Bharat Publications, 1003, Belanganj, Agra, 1956.

(11) TATE, M. W. : 'Statistics in Education' :
New York, Macmillan Co., 1955.

(12) VERNON, P. E. : 'The Measurement of Abilities' ;
University of London Press, Warwick Square, London, E. C. 4, 1940.

(13) WALKER, HELEN M. : 'Elementary Statistical Methods' ;
New York, Henry Holt & Co.

(14) WALKER, HELEN AND LEV, J. : 'Statistical Inference' :
New York, Henry Holt & Co.

OF EDNL. & PSYL. RES
LIBRARY

S.C.E.R.T., West Bengal
Library

Form No. 3.

PSY, RES.L-1

Bureau of Educational & Psychological Research Library.

The book is to be returned within the date stamped last.

WBGP-59/60-5119C-5M